



**Dina Maria Morgado Salvador**

Mestre em Probabilidades e Estatística

**Modelação de Matrizes Estocásticas  
Simétricas  
Operadores do tipo  $\text{vec}$**

Dissertação para obtenção do Grau de Doutor em  
Estatística e Gestão do Risco

Orientador:

Professor Doutor João Tiago Praça Nunes Mexia,  
Professor Catedrático Jubilado,  
Faculdade de Ciências e Tecnologia

**Júri:**

Presidente: Prof. Doutor Jorge Orestes Lasbarrères Cerdeira

Arguente(s): Prof. Doutora Teresa Paula Costa Azinheira Oliveira  
Prof. Doutora Célia Maria Pinto Nunes

Vogais: Prof. Doutor António Gabriel da Silva St. Aubyn  
Prof. Doutor Luís Pedro Carneiro Ramos  
Prof. Doutor Miguel dos Santos Fonseca  
Prof. Doutor João Tiago Praça Nunes Mexia



# Conteúdo

<b>Agradecimentos</b>	<b>xv</b>
<b>Resumo</b>	<b>xvii</b>
<b>Abstract</b>	<b>xxi</b>
<b>Notações e Acrónimos</b>	<b>xxv</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Preliminares</b>	<b>7</b>
2.1 Conceitos e Resultados Básicos . . . . .	7
2.1.1 Matrizes Ortogonais . . . . .	7
2.1.2 Matrizes Simétricas . . . . .	8
2.1.3 Matrizes Inversas Generalizadas e Matrizes Inversas de Moore-Penrose . . . . .	10
2.1.4 Matrizes de Projecção Ortogonal . . . . .	11
2.1.5 Matrizes Ortogonais Estandarizadas e Diagonalizadas . . . . .	17
2.1.6 Produto de Kronecker . . . . .	18
2.1.7 Operador <b>vec</b> . . . . .	22
2.2 Álgebras de Jordan Comutativas . . . . .	25
2.3 Distribuições Associadas à Distribuição Normal . . . . .	35
2.3.1 Distribuição Qui-quadrado Não Central . . . . .	36
2.3.2 Distribuição $F$ de Snedecor Não Central . . . . .	37
2.3.3 Distribuição de Formas Quadráticas . . . . .	38
2.4 Testes de Hipóteses - Testes $F$ . . . . .	38
<b>3 Modelos para matrizes estocásticas simétricas</b>	<b>43</b>
3.1 Considerações Prévias . . . . .	43
3.2 Modelo Base . . . . .	43
3.2.1 Validação do Modelo . . . . .	45
3.2.1.1 Operadores vectoriais . . . . .	45
3.2.1.2 Resultados importantes . . . . .	47

3.2.2	Grau do modelo . . . . .	50
3.2.3	Inferência sobre os parâmetros de estrutura $(\beta_i)$ , $\sigma^2$ e parâmetros de não centralidade $(\delta_r)$ . . . . .	60
3.2.3.1	Inferência sobre $\beta_i$ . . . . .	60
3.2.3.2	Inferência sobre $\sigma^2$ . . . . .	63
3.2.3.3	Inferência sobre $\delta_r$ . . . . .	64
3.2.3.4	Uma aplicação da lineariedade assintótica . . . . .	67
<b>4</b>	<b>Famílias Estruturadas</b>	<b>73</b>
4.1	Considerações Prévias . . . . .	73
4.2	Delineamento Base - Estrutura Algébrica . . . . .	74
4.3	Heterocedasticidade controlada sem restrições . . . . .	74
4.3.1	Análise conjunta de matrizes estocásticas simétricas . . . . .	86
4.3.2	Álgebras de Jordan Comutativas e Cruzamento de Factores . . . . .	88
<b>5</b>	<b>Aplicação Prática</b>	<b>93</b>
5.1	Considerações Prévias . . . . .	93
5.2	Metodologia <i>Statis</i> . . . . .	93
5.2.1	O Método <i>STATIS</i> . . . . .	95
5.2.2	O Método <i>Statis-Dual</i> . . . . .	96
5.2.3	Inter-estrutura . . . . .	98
5.3	Eleições Legislativas em Portugal Continental . . . . .	99
5.3.1	Enquadramento da aplicação prática . . . . .	99
5.3.2	Ajustamento e validação do modelo . . . . .	102
5.4	<i>Semi-Simulação</i> . . . . .	109
<b>6</b>	<b>Conclusão</b>	<b>115</b>

# Lista de Figuras

5.1	Fases da metodologia <i>STATIS</i> : Inter-estrutura e Compromisso. . .	94
5.2	$k$ quadros de dados: notação do método <i>STATIS</i> . . . . .	95
5.3	$k$ quadros de dados: notação do método <i>STATIS-Dual</i> . . . . .	97



# Lista de Tabelas

5.1	Distritos-Concelhos escolhidos por Região . . . . .	101
5.2	As diferentes opções políticas consideradas . . . . .	102
5.3	Valores próprios dos vários concelhos . . . . .	103
5.4	Percentagem de variabilidade explicada nos vários concelhos . . .	104
5.5	Valor da Preponderância para os vários concelhos . . . . .	104
5.6	Vectorios de estrutura dos concelhos da Região Norte . . . . .	105
5.7	Vectorios de estrutura dos concelhos da Região Centro Norte . . .	106
5.8	Vectorios de estrutura dos concelhos da Região Centro . . . . .	106
5.9	Vectorios de estrutura dos concelhos da Região Sul . . . . .	107
5.10	Valor da Estatística $\overline{\mathcal{F}}$ para os vários concelhos . . . . .	107
5.11	Valor da Estatística $\mathcal{F}_1$ e Estimativa de $\delta_1$ para os vários concelhos	108
5.12	Valores próprios das várias matrizes <i>semi-simuladas</i> . . . . .	110
5.13	Percentagem de variabilidade explicada nas várias matrizes <i>semi-simuladas</i> . . . . .	111
5.14	Valor da Estatística $\overline{\mathcal{F}}$ para as várias matrizes <i>semi-simuladas</i> . .	112

5.15	Valor da Estatística $\mathcal{F}_2$ e $\mathcal{F}_1^\circ$ para as várias matrizes <i>semi-simuladas</i>	
	.....	113
5.16	Estimativa de $\delta_2$ para as várias matrizes <i>semi-simuladas</i>	114



Este texto não foi escrito respeitando o Novo Acordo Ortográfico, pois está ao abrigo do período de transição em que as duas ortografias coexistem. O Artigo 2.º da Resolução da Assembleia da República n.º 35/2008, publicada no Diário da República, 1.ª série, n.º 145, de 29 de Julho de 2008, pp. 4802-4803, prevê um prazo de seis anos para a transição definitiva para a nova ortografia. Assim, este período, iniciado a 13 de Maio de 2009, para a entrada em vigor do novo Acordo Ortográfico, terminará em 2015.

x

.

## Modelação de Matrizes Estocásticas Simétricas - - Operadores do tipo vec

### **Direitos de cópia**

**Copyright© 2013** – Todos os direitos reservados. Dina Maria Morgado Salvador. Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

### **Copyright**

**Copyright© 2013** – All rights reserved. Dina Maria Morgado Salvador. Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Faculdade de Ciências e Tecnologia and Universidade Nova de Lisboa have the perpetual right with no geographical boundaries to archive and publish this dissertation through printed copies reproduced on paper or digital form or by any means known or to be invented, and to divulge through scientific repositories and admit your copy and distribution for educational purposes or research, not commercial, as long as the credit is given to the author and editor.



*Aos meus Pais,*

*que por uma vida de dedicação, amor e trabalho sempre possibilitaram aos seus filhos a oportunidade de realizar sonhos e conquistas.*



# Agradecimentos

O presente trabalho de tese de doutoramento dependeu do contributo de várias formas e diferentes níveis por parte de várias pessoas, pois envolveu um grande conjunto de apoios humanos e materiais, que foram fundamentais para a sua execução. Embora se trate de um trabalho individual, gostaria de expressar o meu reconhecimento e agradecimento a todos aqueles que me ajudaram ao longo deste trabalho.

Em primeiro lugar quero agradecer ao meu orientador. A conclusão desta dissertação não teria sido possível sem a ajuda, conselhos, disponibilidade e paciência do meu orientador, o Professor Doutor João Tiago Mexia. Sou muito grata a ele pela confiança, persistência e suporte, e que com a sua competência teórica sempre auxiliou no delineamento e sistematização do trabalho em todas as etapas desta longa caminhada.

Agradeço também à amiga e companheira de doutoramento (muitas vezes designada "companheira de desgraça") Sandra Monteiro, que comigo partilhou as angústias e sofrimento decorrentes de um trabalho desta natureza. A ela agradeço não só a amizade, carinho e auxílio, mas também o sorriso e bom humor constantes, que em muitas horas foram necessários.

Não posso esquecer, as amigas de todas as horas, Catarina e Anabela, que sempre incentivaram e motivaram nesta caminhada.

Quero agradecer também a alguns colegas do Instituto Politécnico de Setúbal e a todos que demonstraram interesse em saber "Como vai a tese?" ou que faziam a dolorosa pergunta "Já terminaste?", e que tornaram menos pesada esta tarefa criando motivação para concluir este longo trabalho, tornando possível a resposta favorável e desejada a tais perguntas.

Por fim, mas não menos importante, quero agradecer o insubstituível apoio dos meus pais, do meu irmão Paulo, da minha cunhada Catarina, do João e da minha querida Mali, bem como da restante família, que ao pensarem que pouco estavam a contribuir, em muito ajudaram com o enorme carinho, apoio, amor, amizade e companheirismo manifestado em todos os momentos, em especial naqueles momentos mais difíceis e em que o desânimo parecia querer instalar-se.

A todos que, directa ou indirectamente contribuíram para a realização deste

trabalho, e a todos aqueles que neste instante, traída pela memória, não recordo e/ou não menciono, **o meu sincero reconhecimento e muito obrigada!**

*Este trabalho de Investigação foi feito ao abrigo do Programa de Apoio à Formação Avançada de Docentes do Ensino Superior Politécnico (PROTEC) do Instituto Politécnico de Setúbal.*

*A elaboração desta tese beneficiou do regime de isenção de propinas de doutoramento, no âmbito do Protocolo de Cooperação existente entre a Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa e o Instituto Politécnico de Setúbal.*



# Resumo

Considere-se uma matriz de dados cujas linhas correspondem aos objectos e as colunas correspondem às variáveis. A estas matrizes podemos associar-lhes matrizes diagonais de pesos, quer para os objectos, quer para as variáveis, obtendo-se assim um trio de matrizes a que se dá o nome de estudo. Ao conjunto de vários estudos, chama-se série de estudos.

Hoje em dia, existem várias técnicas para o tratamento de séries de estudos na Análise Conjunta de Tabelas. A formulação apresentada permite-nos fazer inferência sobre as séries de estudo, pois os resultados apresentados nesta dissertação podem ser aplicados às matrizes dos produtos de Hilbert-Schmidt, matrizes estas muito importantes na primeira fase da metodologia *Statis*.

Os modelos que consideramos são modelos que se baseiam na decomposição espectral das matrizes médias e que são apresentados da seguinte forma

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} = \sum_{i=1}^r \beta_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}}$$

para matrizes com grau  $r \geq 1$ , onde  $\bar{\mathbf{E}} = \frac{1}{2} (\mathbf{E} + \mathbf{E}^t)$  com

$$\mathbf{vec}(\mathbf{E}) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right),$$

como podemos ver em [Salvador *et al.*, 2012].

Para estudar tais modelos, iremos usar operadores do tipo **vec**, ver [Salvador e Mexia, 2013]. Recorde-se que os operadores **vec** são operadores que quando aplicados a uma matriz, empilham sucessivamente as colunas dessa matriz, ver [Harville, 1998] e [Schott, 1997]. Estes operadores do tipo **vec** são operadores que também fazem corresponder vectores a matrizes.

Assim, para uma matriz simétrica  $\mathbf{A} = [a_{ij}]$  de ordem  $k$ , temos os operadores

$$s\_vec(\mathbf{A}) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{rr} \\ \vdots \\ a_{kk} \\ \sqrt{2}a_{12} \\ \vdots \\ \sqrt{2}a_{(k-1)k} \end{bmatrix},$$

$$a\_vec(\mathbf{A};r) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{rr} \\ \sqrt{2}a_{12} \\ \vdots \\ \sqrt{2}a_{rk} \end{bmatrix} \quad \text{e} \quad b\_vec(\mathbf{A};r) = \begin{bmatrix} a_{(r+1)(r+1)} \\ \vdots \\ a_{kk} \\ \sqrt{2}a_{(r+1)(r+2)} \\ \vdots \\ \sqrt{2}a_{(k-1)k} \end{bmatrix}$$

onde

- $s\_vec(\mathbf{A})$  é um vector com  $g$  componentes, em que  $g = \frac{k(k+1)}{2}$ ;
- $a\_vec(\mathbf{A};r)$  é um vector com  $a_r$  componentes, em que  $a_r = g - b_r$ ;
- $b\_vec(\mathbf{A};r)$  é um vector com  $b_r$  componentes, em que  $b_r = \frac{(k-r)(k-r+1)}{2}$ .

Temos também

$$\mathbf{v}_j(\mathbf{A}) = \begin{bmatrix} a_{jj} \\ a_{j(j+1)} \\ \vdots \\ a_{j(k-1)} \\ \vdots \\ a_{jk} \end{bmatrix}$$

que é um vector coluna cujas componentes são os elementos do  $j$ -ésimo vector linha da matriz  $\mathbf{A}$  que começa na posição  $(j, j)$ . O vector  $\mathbf{v}_j(\mathbf{A})$  é um vector com  $c_j$  componentes, em que  $c_j = k - j + 1$ .

Aplicando estes operadores à matriz  $\mathbf{S}$ , mostrou-se que se  $r \leq k$ , as componentes da  $b\_vec(\mathbf{P}^*\mathbf{S}(\mathbf{P}^*)^t; r)$  têm uma distribuição aproximadamente normal, ou seja

$$b\_vec(\mathbf{P}^*\mathbf{S}(\mathbf{P}^*)^t; r) \rightsquigarrow N(0, \sigma^2),$$

sendo  $\mathbf{P}^* = [\gamma_1^t \cdots \gamma_k^t]^t$  uma matriz ortogonal, em que  $\gamma_1, \dots, \gamma_k$  são os vectores próprios associados aos valores próprios  $\theta_1, \dots, \theta_k$  da matriz  $\mathbf{S}$  cuja decomposição espectral é

$$\mathbf{S} = \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t.$$

Com vista a encontrar o grau do modelo, temos a estatística de teste

$$\mathcal{F}_r^\circ = \frac{b_r}{k-r+1} \frac{\|\underline{\mathbf{v}}_r (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t)\|^2}{\|b_- \mathbf{vec}(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t; r)\|^2}$$

que tem aproximadamente uma distribuição  $F$  de Snedecor não central, com  $k-r+1$  e  $b_r$  graus de liberdade, ver [Mexia, 1989a] e [Salvador e Mexia, 2013], ou seja,

$$\mathcal{F}_r^\circ \rightsquigarrow F_{k-r+1, b_r}(\delta_r)$$

com

$$\delta_r = \frac{\lambda_r^2}{\sigma^2},$$

e que permite testar a hipótese  $H_{0,r}^\circ : \lambda_r = 0$ .

Usando estes resultados é possível determinar  $r$  e ajustar um modelo do tipo

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \overline{\mathbf{E}}.$$

Este modelo pode ser simplificado se o valor de  $\lambda_r$  for igual a 0, reduzindo o modelo anterior ao modelo

$$\mathbf{S} = \sum_{i=1}^{r-1} \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \overline{\mathbf{E}}.$$

Caso  $\lambda_{r-1}$  seja significativamente diferente de 0 então este último modelo será um modelo com grau  $r-1$ .

**Palavras chave:** Matrizes Simétricas, Modelos para Matrizes Estocásticas Simétricas, Operadores do tipo **vec** e Famílias Estruturadas.



# Abstract

Consider a matrix of data whose rows correspond to objects and whose columns correspond to variables. We can associate them to diagonal weights matrices, one for the objects and the other for the variables, thus obtaining a trio of arrays to which is given the name of study. A set of several studies, called a series of studies.

Today several techniques exist for treating a series of studies on the Joint Analysis of Tables. The presented formulation allows us to make inference about the series of studies, since the results presented in this thesis can be applied to matrices of Hilbert-Schmidt products that are very important in first step of the *Statis* methodology.

The models we consider are based in the spectral decomposition of the mean matrices being of the form

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} = \sum_{i=1}^r \beta_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}}$$

for matrices with degree  $r \geq 1$ , where  $\bar{\mathbf{E}} = \frac{1}{2}(\mathbf{E} + \mathbf{E}^t)$  with

$$\mathbf{vec}(\mathbf{E}) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right),$$

as we see in [Salvador *et al.*, 2012].

To study such models we used **vec** type operators, see [Salvador e Mexia, 2013]. Remember that **vec** operators are operators that when applied to a matrix, successively stacking the columns of this matrix, see [Harville, 1998] and [Schott, 1997]. The **vec** type operators are operators are also vectors correspond to matrices.

Namely for a symmetric matrix  $\mathbf{A} = [a_{ij}]$  of order  $k$ , we have

$$s\_vec(\mathbf{A}) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{rr} \\ \vdots \\ a_{kk} \\ \sqrt{2}a_{12} \\ \vdots \\ \sqrt{2}a_{(k-1)k} \end{bmatrix},$$

$$a\_vec(\mathbf{A};r) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{rr} \\ \sqrt{2}a_{12} \\ \vdots \\ \sqrt{2}a_{rk} \end{bmatrix} \quad \text{and} \quad b\_vec(\mathbf{A};r) = \begin{bmatrix} a_{(r+1)(r+1)} \\ \vdots \\ a_{kk} \\ \sqrt{2}a_{(r+1)(r+2)} \\ \vdots \\ \sqrt{2}a_{(k-1)k} \end{bmatrix}$$

where

- $s\_vec(\mathbf{A})$  is a vector with  $g$  components where  $g = \frac{k(k+1)}{2}$ ;
- $a\_vec(\mathbf{A};r)$  is a vector with  $a_r$  components where  $a_r = g - b_r$ ;
- $b\_vec(\mathbf{A};r)$  is a vector with  $b_r$  components where  $b_r = \frac{(k-r)(k-r+1)}{2}$ .

We have also

$$\mathbf{v}_j(\mathbf{A}) = \begin{bmatrix} a_{jj} \\ a_{j(j+1)} \\ \vdots \\ a_{j(k-1)} \\ \vdots \\ a_{jk} \end{bmatrix}$$

which is a column vector whose components are the components of the  $j$ -th row vector of matrix  $\mathbf{A}$  starting at position  $(j, j)$ . The vector  $\mathbf{v}_j(\mathbf{A})$  have  $c_j$  components where  $c_j = k - j + 1$ .

Applying these operators to  $\mathbf{S}$ , we showed that if  $r \leq k$ , the components of  $b\_vec(\mathbf{P}^*\mathbf{S}(\mathbf{P}^*)^t; r)$  have approximately normal distribution, namely

$$b\_vec(\mathbf{P}^*\mathbf{S}(\mathbf{P}^*)^t; r) \rightsquigarrow N(0, \sigma^2),$$

where  $\mathbf{P}^* = [\gamma_1^t \cdots \gamma_k^t]^t$  is an orthogonal matrix where  $\gamma_1, \dots, \gamma_k$  are the eigenvectors associated with eigenvalues  $\theta_1, \dots, \theta_k$  of matrix  $\mathbf{S}$  whose spectral decomposition is

$$\mathbf{S} = \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t.$$

In order to find the model degree, we have the statistic

$$\mathcal{F}_r^\circ = \frac{b_r}{k-r+1} \frac{\|\underline{\mathbf{v}}_r (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t)\|^2}{\|b_- \mathbf{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t; r)\|^2}$$

which has approximately a non-central  $F$  distribution, with  $k-r+1$  and  $b_r$  degrees of freedom, see [Mexia, 1989a] and [Salvador e Mexia, 2013], that is

$$\mathcal{F}_r^\circ \rightsquigarrow F_{k-r+1, b_r}(\delta_r)$$

where

$$\delta_r = \frac{\lambda_r^2}{\sigma^2},$$

so that we may test  $H_{0,r}^\circ : \lambda_r = 0$ .

Using these results it is possible to determine  $r$  and fit a model

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \overline{\mathbf{E}}.$$

This model can be simplified if the value of  $\lambda_r$  is equal to 0, reducing the previous model to model

$$\mathbf{S} = \sum_{i=1}^{r-1} \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \overline{\mathbf{E}}.$$

If  $\lambda_{r-1}$  is significantly different from 0 then the last model is a model with degree  $r-1$ .

**Keywords :** Symmetric Matrices, Models for Symmetric Stochastic Matrices, **vec** type Operators and Structured Families.





# Notações e Acrónimos

$\mathbf{v}$	$\rightarrow$	vector (letra minúscula a negrito)
$\mathbf{0}^n$	$\rightarrow$	vector nulo do tipo $n \times 1$
$\mathbf{1}^n$	$\rightarrow$	vector com as componentes todas iguais a 1 do tipo $n \times 1$
$\mathbf{Y}$	$\rightarrow$	vector aleatório (letra maiúscula a negrito)
$\mathbf{A}$	$\rightarrow$	matriz aleatória (letra maiúscula a negrito)
$\mathbf{O}_{n,m}$	$\rightarrow$	matriz nula do tipo $n \times m$
$\mathbf{I}_n$	$\rightarrow$	matriz identidade de ordem $n$
$\mathbf{J}_n$	$\rightarrow$	matriz de ordem $n$ com as componentes todas iguais a 1
$\mathbf{A}^t$	$\rightarrow$	matriz transposta de $\mathbf{A}$
$\mathbf{A}^{-1}$	$\rightarrow$	matriz inversa de $\mathbf{A}$
$\mathbf{A}^+$	$\rightarrow$	matriz inversa de Moore-Penrose de $\mathbf{A}$
$\mathbf{A}^-$	$\rightarrow$	matriz inversa generalizada de $\mathbf{A}$
$\mathbf{A} = [a_{ij}]$	$\rightarrow$	matriz $\mathbf{A}$ com elemento genérico $a_{ij}$
$\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_k]$	$\rightarrow$	matriz com vectores coluna $\mathbf{a}_1, \dots, \mathbf{a}_k$
$\mathbf{D}(a_1, \dots, a_n)$	$\rightarrow$	matriz diagonal com elementos principais $a_1, \dots, a_n$
$\ \cdot\ $	$\rightarrow$	norma euclidiana
$\text{car}(\mathbf{A})$	$\rightarrow$	característica da matriz $\mathbf{A}$
$ \mathbf{A} $	$\rightarrow$	determinante da matriz $\mathbf{A}$

$tr(\mathbf{A})$	$\rightarrow$	traço da matriz $\mathbf{A}$
$\mathcal{R}(\mathbf{A})$	$\rightarrow$	espaço imagem da matriz $\mathbf{A}$
$\mathcal{N}(\mathbf{A})$	$\rightarrow$	espaço nulidade da matriz $\mathbf{A}$
$\mathcal{Q}(\mathbf{A})$	$\rightarrow$	matriz de projecção ortogonal sobre o espaço imagem da matriz $\mathbf{A}$
$P_S$	$\rightarrow$	matriz de projecção ortogonal sobre o subespaço $S$
$\dim(S)$	$\rightarrow$	dimensão do subespaço $S$
$S^\perp$	$\rightarrow$	complemento ortogonal de $S$
$S_E^\perp$	$\rightarrow$	complemento ortogonal do subespaço $S$ relativamente a $E$
$\boxplus$	$\rightarrow$	soma directa ortogonal de subespaços
$E[X]$	$\rightarrow$	valor esperado da variável aleatória $X$
$V[X]$	$\rightarrow$	variância da variável aleatória $X$
$E[\mathbf{X}]$	$\rightarrow$	valor esperado do vector aleatório $\mathbf{X}$
$\mathbf{COV}[\mathbf{X}]$	$\rightarrow$	matriz de covariância do vector aleatório $\mathbf{X}$
$\Pr()$	$\rightarrow$	probabilidade
$\delta_i^k$	$\rightarrow$	vector coluna do tipo $k \times 1$ cuja $i$ – ésima componente é igual a 1 e as restantes nulas
$\otimes$	$\rightarrow$	produto de Kronecker entre matrizes
$\rightsquigarrow$	$\rightarrow$	...distribui-se aproximadamente como...
$\approx$	$\rightarrow$	aproximadamente igual
$\sim$	$\rightarrow$	...distribui-se como...
$\longrightarrow$	$\rightarrow$	...converge para...
$\xrightarrow{q.c.}$	$\rightarrow$	...converge quase certamente para...
$\xrightarrow{u}$	$\rightarrow$	...converge uniformemente para...

$\hookrightarrow$	$\rightarrow$	distribuição limite
$N(\mu, \sigma^2)$	$\rightarrow$	distribuição normal com valor médio $\mu$ e variância $\sigma^2$
$\chi_n^2$	$\rightarrow$	distribuição qui-quadrado central com $n$ graus de liberdade
$\chi_n^2(\delta)$	$\rightarrow$	distribuição qui-quadrado com $n$ graus de liberdade e parâmetro de não centralidade $\delta$
$F_{m,n}$	$\rightarrow$	distribuição $F$ de Snedecor central com $m$ e $n$ graus de liberdade
$F_{m,n}(\delta)$	$\rightarrow$	distribuição $F$ de Snedecor com $m$ e $n$ graus de liberdade e parâmetro de não centralidade $\delta$
$\overline{F}_{m,n}$	$\rightarrow$	distribuição do quociente de dois qui-quadrados centrais independentes, $\chi_m^2$ e $\chi_n^2$
$\overline{F}_{m,n}(\delta)$	$\rightarrow$	distribuição quociente de dois qui-quadrados independentes, $\chi_m^2(\delta)$ e $\chi_n^2$
$\overline{f}(z m, n)$	$\rightarrow$	função densidade de uma distribuição $\overline{F}_{m,n}$
$\overline{f}(z m, n, \delta)$	$\rightarrow$	função densidade de uma distribuição $\overline{F}_{m,n}(\delta)$
■	$\rightarrow$	final de demonstração
STATIS	$\rightarrow$	<i>Structuration de Tableaux à Trois Indices de la Statistique</i>
ACT	$\rightarrow$	Análise Conjunta de Tabelas
<i>i.i.d.</i>	$\rightarrow$	independentes e identicamente distribuídos
ANOVA	$\rightarrow$	Análise de Variância



# Capítulo 1

## Introdução

No âmbito da tese de doutoramento em Estatística e Gestão do Risco, realizada na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, o presente trabalho pretende contribuir para o estudo da estrutura algébrica de modelos.

Nesta dissertação pretendemos introduzir uma nova formulação para os modelos de matrizes estocásticas simétricas. Nessa nova formulação, iremos utilizar os operadores  $s\_vec(\mathbf{A})$ ,  $a\_vec(\mathbf{A};r)$  e  $b\_vec(\mathbf{A};r)$  aplicados a uma certa matriz  $\mathbf{A}$ , que serão abordados na subsecção 3.2.1.1 e utilizados na apresentação de resultados importantes para construir modelos para matrizes estocásticas simétricas. Estes operadores vectoriais, além de se apresentarem como parte importante na nova formulação para estes modelos, também facilitam a apresentação desses resultados.

Os modelos para matrizes estocásticas simétricas deram a base para realizar inferência para matrizes isoladas e para famílias estruturadas de matrizes. Nestas famílias as matrizes, todas da mesma ordem, correspondem aos tratamentos de modelos base, conforme apresentado na secção 3.2.

Uma vez que as matrizes têm todas a mesma ordem, estamos perante o caso equilibrado em que temos o mesmo número de graus de liberdade para o erro para cada tratamento. A ANOVA e técnicas relacionadas são, no caso equilibrado, técnicas robustas para o caso de heterocedasticidade e ainda mais para o caso da não-normalidade, ver [Ito, 1980] e [Scheffé, 1959]. Recorde-se que, para modelos individuais de matrizes simétricas, uma formulação interessante é apresentada em [Areia, 2009].

Admitindo que a série de estudos tem uma estrutura comum de grau  $r < k$  e portanto, em vez de considerarmos o modelo

$$\mathbf{S} = \lambda \boldsymbol{\alpha} \boldsymbol{\alpha}^t + \overline{\mathbf{E}} \quad (1.1)$$

sendo  $\mathbf{S}$  a matriz dos produtos internos de Hilbert-Schmidt, apresentado em

[Oliveira e Mexia, 1999], iremos considerar o modelo da seguinte forma

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} = \sum_{i=1}^r \boldsymbol{\beta}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} \quad (1.2)$$

com  $\boldsymbol{\alpha}_i^t \boldsymbol{\alpha}_j = \delta_{ij}$ ,  $\boldsymbol{\beta}_i = \lambda_i \boldsymbol{\alpha}_i$ , com  $i, j = 1, \dots, r$  em que  $\delta_{ij} = 0$  se  $i \neq j$  e  $\delta_{ij} = 1$  se  $i = j$ . De forma que a matriz  $\mathbf{S}$  seja simétrica, admite-se  $\bar{\mathbf{E}} = \frac{1}{2} (\mathbf{E} + \mathbf{E}^t)$  com

$$\text{vec}(\mathbf{E}) \sim N(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}). \quad (1.3)$$

Em [Oliveira, 2001], através de um estudo de simulação, mostrou-se que quando o primeiro valor próprio da matriz  $\mathbf{S}$  é suficientemente maior que os restantes, pode-se utilizá-lo para estimar  $\lambda_1$  e o correspondente vector próprio para estimar  $\boldsymbol{\alpha}_1$ . Pretende-se generalizar esta abordagem, ou seja, quando os  $r$  primeiros valores próprios da matriz  $\mathbf{S}$  são suficientemente maiores que os restantes, pode-se utilizá-los para estimar  $\lambda_1, \dots, \lambda_r$  e os correspondentes vectores próprios para estimar  $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_r$ .

Assim, sejam  $\theta_1 > \theta_2 > \dots > \theta_k$  os valores próprios da matriz  $\mathbf{S}$  associados aos vectores próprios  $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_k$ , respectivamente. Assim

$$\mathbf{S} = \sum_{i=1}^k \theta_i \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^t. \quad (1.4)$$

Então podem estimar-se os  $r$  primeiros vectores estrutura por

$$\tilde{\boldsymbol{\beta}}_i = \theta_i \boldsymbol{\gamma}_i, \text{ com } i = 1, \dots, r. \quad (1.5)$$

Os casos mais interessantes para estes tipos de modelos, ocorrem quando as matrizes dessa família têm o primeiro valor próprio dominante ou os primeiros valores próprios dominantes. Nesse caso, estuda-se a acção dos factores desse delineamento, sobre os componentes do primeiro vector de estrutura ou dos primeiros vectores de estrutura, conforme seja apenas o primeiro valor próprio dominante ou os primeiros valores próprios dominantes, respectivamente.

Assim dada uma matriz simétrica  $\mathbf{A} = [a_{ij}]$  de ordem  $k$ , ou seja  $\mathbf{A} \in \Theta_k$  em que  $\Theta_k$  representa o conjunto de matrizes simétricas de ordem  $k$ , tal que

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1(r-1)} & a_{1r} & a_{1(r+1)} & \cdots & a_{1(k-1)} & a_{1k} \\ a_{12} & a_{22} & \cdots & a_{2(r-1)} & a_{2r} & a_{2(r+1)} & \cdots & a_{2(k-1)} & a_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{1(r-1)} & a_{2(r-1)} & \cdots & a_{(r-1)(r-1)} & a_{(r-1)r} & a_{(r-1)(r+1)} & \cdots & a_{(r-1)(k-1)} & a_{(r-1)k} \\ a_{1r} & a_{2r} & \cdots & a_{(r-1)r} & a_{rr} & a_{r(r+1)} & \cdots & a_{r(k-1)} & a_{rk} \\ a_{1(r+1)} & a_{2(r+1)} & \cdots & a_{(r-1)(r+1)} & a_{r(r+1)} & a_{(r+1)(r+1)} & \cdots & a_{(r+1)(k-1)} & a_{(r+1)k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{1(k-1)} & a_{2(k-1)} & \cdots & a_{(r-1)(k-1)} & a_{r(k-1)} & a_{(r+1)(k-1)} & \cdots & a_{(k-1)(k-1)} & a_{(k-1)k} \\ a_{1k} & a_{2k} & \cdots & a_{(r-1)k} & a_{rk} & a_{(r+1)k} & \cdots & a_{(k-1)k} & a_{kk} \end{bmatrix},$$

representaremos por  $s\_vec(\mathbf{A})$  o vector coluna

$$s\_vec(\mathbf{A}) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{rr} \\ \vdots \\ a_{kk} \\ \sqrt{2}a_{12} \\ \vdots \\ \sqrt{2}a_{(k-1)k} \end{bmatrix} \quad (1.6)$$

do tipo  $g \times 1$  com  $g = \frac{k(k+1)}{2}$ , tendo-se

$$\|\mathbf{A}\|^2 = \|s\_vec(\mathbf{A})\|^2. \quad (1.7)$$

Consideremos ainda mais dois operadores,  $a\_vec(\mathbf{A}; r)$  e  $b\_vec(\mathbf{A}; r)$ , dados por

$$a\_vec(\mathbf{A}; r) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{rr} \\ \sqrt{2}a_{12} \\ \vdots \\ \sqrt{2}a_{rk} \end{bmatrix}; \quad b\_vec(\mathbf{A}; r) = \begin{bmatrix} a_{(r+1)(r+1)} \\ \vdots \\ a_{kk} \\ \sqrt{2}a_{(r+1)(r+2)} \\ \vdots \\ \sqrt{2}a_{(k-1)k} \end{bmatrix} \quad (1.8)$$

do tipo  $a_r \times 1$  e  $b_r \times 1$ , respectivamente, com  $a_r = g - b_r$  e  $b_r = \frac{(k-r)(k-r+1)}{2}$ .

Consideremos também o vector coluna

$$\mathbf{v}_j(\mathbf{A}) = \begin{bmatrix} a_{jj} \\ a_{j(j+1)} \\ \vdots \\ a_{j(k-1)} \\ \vdots \\ a_{jk} \end{bmatrix} \quad (1.9)$$

cujas componentes são os elementos do  $j$ -ésimo vector linha da matriz  $\mathbf{A}$  que começa na posição  $(j, j)$ , portanto será um vector coluna do tipo  $c_j \times 1$ , em que  $c_j = k - j + 1$ .

Aplicando estes operadores à matriz  $\mathbf{S}$ , mostrou-se que se  $r \leq k$ , as componentes de  $b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t; r)$  têm uma distribuição aproximadamente normal, ou seja

$$b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t; r) \rightsquigarrow N(0, \sigma^2), \quad (1.10)$$

sendo  $\mathbf{P}^* = [\gamma_1^t \ \cdots \ \gamma_k^t]^t$  uma matriz ortogonal, em que  $\gamma_1, \dots, \gamma_k$  são os vectores próprios associados aos valores próprios  $\theta_1, \dots, \theta_k$  da matriz  $\mathbf{S}$  cuja decomposição espectral é

$$\mathbf{S} = \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t.$$

Com vista a encontrar o grau do modelo, a estatística de teste usada para o validar é

$$\mathcal{F}_r^\circ = \frac{b_r}{k - r + 1} \frac{\|\mathbf{v}_r(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t)\|^2}{\|b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t; r)\|^2} \quad (1.11)$$

que tem aproximadamente uma distribuição  $F$  de Snedecor não central, com  $k - r + 1$  e  $b_r$  graus de liberdade, ver [Mexia, 1989a] e [Salvador e Mexia, 2013], ou seja,

$$\mathcal{F}_r^\circ \rightsquigarrow F_{k-r+1, b_r}(\delta_r) \quad (1.12)$$

com

$$\delta_r = \frac{\lambda_r^2}{\sigma^2}. \quad (1.13)$$

Usando estes resultados é possível determinar  $r$  e ajustar um modelo do tipo (1.2).

De facto, um ponto de interesse abordado nesta dissertação, diz respeito à variação do grau da estrutura comum, ou seja, a variação do grau  $r$ . Sabemos que quando se considera um valor para  $r$  inferior ao que deveria ser considerado, isto é, quando o  $r$  é sub-avaliado, a soma dos quadrados dos erros terá um parâmetro de



não centralidade bastante grande e neste caso a potência dos testes  $F$  diminuirá, ver [Mexia, 1989a].

Note-se ainda que o modelo pode ser simplificado se o valor de  $\lambda_r$  for igual a 0, reduzindo-se o modelo anterior ao modelo

$$\mathbf{S} = \sum_{i=1}^{r-1} \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}}. \quad (1.14)$$

Caso  $\lambda_{r-1}$  seja significativamente diferente de 0 então este último modelo será um modelo com grau  $r - 1$ .

Uma vez apresentado o tema em que se insere esta dissertação, vejamos agora como esta está estruturada.

No capítulo 2 começa-se por apresentar os resultados preliminares de forma a ter-se uma dissertação auto-suficiente. Estes resultados serão de natureza algébrica e de natureza estatística. Nele incluiu-se alguns conceitos e resultados muito importantes sobre teoria das matrizes com aplicação à estatística, ver [Schott, 1997], [Horn e Johnson, 1985] e [Rao, 1973].

De seguida, no capítulo 3, que é um capítulo que constitui a parte essencial da nossa dissertação, apresentaremos os modelos para matrizes estocásticas simétricas baseados na decomposição espectral das respectivas matrizes médias, de acordo com uma nova formulação, introduzindo os operadores do tipo **vec**.

No capítulo seguinte, capítulo 4, apresentaremos as famílias estruturadas de matrizes estocásticas simétricas, sendo este um capítulo também essencial nesta dissertação. As matrizes de uma tal família terão a mesma dimensão, sendo no nosso caso  $k \times k$ , e corresponderão aos  $n$  tratamentos de um delineamento base. Ao estudar-se uma tal família procura-se analisar a acção dos factores do delineamento base sobre os modelos da família. Os delineamentos que se consideram são delineamentos ortogonais estando associados a partições ortogonais de  $\mathbb{R}^n$ . Assim, como veremos, podemos aplicar a ANOVA e técnicas associadas na análise.

Admitindo-se que os graus  $r_1, \dots, r_n$  dos modelos da família não são nulos, ter-se-á

$$r = \min \{r_1, \dots, r_n\}$$

com  $r > 0$ , podendo estudar-se a acção dos factores do delineamento base sobre os  $r$  primeiros vectores de estrutura.

Quando o grau  $r$  é o mesmo para todas as matrizes da família, estando-se no caso equilibrado, as mesmas contribuem com o mesmo número de graus de liberdade para o erro. Estamos assim no caso equilibrado e poderá admitir-se a homocedasticidade. Assim, ver [Scheffé, 1959] e [Ito, 1980], no caso equilibrado, a ANOVA e técnicas relacionadas são robustas para a heterocedasticidade.

No capítulo 5 incluiu-se uma aplicação prática dos resultados apresentados e desenvolvidos nos capítulos anteriores. Os dados utilizados dizem respeito aos resultados de 12 eleições legislativas realizadas em Portugal Continental entre os anos 1976 e 2009, em 16 concelhos de Portugal Continental. A aplicação prática desta dissertação será apresentada no âmbito da metodologia *Statis* pelo que neste mesmo capítulo foi feito o enquadramento dessa metodologia.

As principais conclusões obtidas no decorrer deste trabalho serão descritas no capítulo 6. Ainda neste capítulo serão também referidos os trabalhos que poderão ser realizados no futuro.

Será ainda de referir que ao longo da elaboração desta dissertação resultaram algumas publicações e apresentações em conferências, como sejam [Salvador *et al.*, 2012], [Salvador *et al.*, 2012a] e [Salvador e Mexia, 2013], e aguarda-se a aceitação de mais uma publicação.

# Capítulo 2

## Preliminares

### 2.1 Conceitos e Resultados Básicos

Este capítulo contém uma compilação de alguns conceitos e resultados muito importantes e bem conhecidos sobre teoria das matrizes com aplicação à estatística. Neste mesmo capítulo, apresentaremos também outros resultados que derivam directamente dos anteriores.

A Álgebra Matricial desempenha um papel importante na Estatística. Em particular, na formulação de modelos lineares é usual usar álgebra matricial na apresentação e verificação dos resultados, porque nos permite manipular de forma eficiente a complexidade das várias variáveis observadas.

Assim, dada a importância dos conceitos e resultados que apresentaremos, estes servirão de apoio ao longo deste trabalho, pois pretende-se que esta dissertação seja auto-suficiente, daí considerar-se pertinente a sua inclusão.

Alguns destes resultados podem ser encontrados, por exemplo, em [Schott, 1997], [Horn e Johnson, 1985] e em [Rao, 1973], que são consideradas referências bibliográficas de Álgebra Matricial para a Estatística. A maior parte das demonstrações dos resultados foram excluídas, uma vez que não foram consideradas de maior importância na elaboração desta dissertação e poderão ser encontradas nas referências citadas.

#### 2.1.1 Matrizes Ortogonais

**Definição 2.1.1 (*Matriz Ortogonal*)** Uma matriz  $\mathbf{P}$  de ordem  $m$ , cujas colunas (ou linhas) formam um conjunto ortonormado de vectores é designada por uma matriz ortogonal, ou seja,  $\mathbf{P}$  é ortogonal se e só se

$$\mathbf{P}^t \mathbf{P} = \mathbf{P} \mathbf{P}^t = \mathbf{I}_m,$$

onde  $\mathbf{I}_m$  representa a matriz identidade de ordem  $m$ .

Vejamos agora algumas propriedades das matrizes ortogonais.

**Teorema 2.1.1** *Sejam  $\mathbf{P}$  e  $\mathbf{Q}$  matrizes ortogonais de ordem  $m$  e  $\mathbf{A}$  uma matriz qualquer também de ordem  $m$ . Então*

- $|\mathbf{P}| = \pm 1$
- $|\mathbf{P}^t \mathbf{A} \mathbf{P}| = |\mathbf{A}|$
- $\mathbf{P} \mathbf{Q}$  é uma matriz ortogonal

onde  $|\cdot|$  representa o determinante de uma matriz.

## 2.1.2 Matrizes Simétricas

Se a transposta de uma matriz  $\mathbf{A}$  é igual à matriz  $\mathbf{A}$ , dizemos que a matriz  $\mathbf{A}$  é simétrica.

**Teorema 2.1.2** *Seja  $\mathbf{A}$  uma matriz simétrica de ordem  $n$ .*

- *Os valores próprios de  $\mathbf{A}$ ,  $\theta_1, \dots, \theta_n$ , são números reais.*
- *Os vectores próprios,  $\gamma_1, \dots, \gamma_k$ , correspondentes aos valores próprios distintos,  $\theta_1, \dots, \theta_k$ , são mutuamente ortogonais. Os vectores próprios,  $\gamma_{k+1}, \dots, \gamma_n$ , correspondentes aos valores próprios não distintos podem ser escolhidos para ser mutuamente ortogonais uns aos outros.*

Se os vectores próprios de uma matriz simétrica  $\mathbf{A}$  são vectores normalizados e são as colunas de uma matriz  $\mathbf{P}$ , então  $\mathbf{P}$  é uma matriz ortogonal que pode ser usada para expressar  $\mathbf{A}$  em termos de seus valores e vectores próprios.

**Teorema 2.1.3** *Se  $\mathbf{A}$  é uma matriz simétrica com valores próprios  $\theta_1, \dots, \theta_n$ , e vectores próprios normalizados  $\gamma_1, \dots, \gamma_n$  então  $\mathbf{A}$  pode ser expressa como*

$$\mathbf{A} = \mathbf{P} \mathbf{D}(\theta_1, \dots, \theta_n) \mathbf{P}^t = \sum_{i=1}^n \theta_i \gamma_i \gamma_i^t$$

onde  $\mathbf{P}$  é uma matriz cujas colunas são os vectores normalizados  $\gamma_1, \dots, \gamma_n$ . A  $\sum_{i=1}^n \theta_i \gamma_i \gamma_i^t$  chama-se decomposição espectral de  $\mathbf{A}$ .

**Dem.** Pelo teorema 2.1.2, sendo  $\mathbf{P}$  uma matriz cujas colunas são os vectores normalizados  $\gamma_1, \dots, \gamma_n$ , então  $\mathbf{P}$  é uma matriz ortogonal. Então temos

$$\begin{aligned}
 \mathbf{P}\mathbf{P}^t &= \mathbf{I}_n \Leftrightarrow \mathbf{A}\mathbf{P}\mathbf{P}^t = \mathbf{A} \Leftrightarrow \mathbf{A} \begin{bmatrix} \gamma_1 & \dots & \gamma_n \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix}^t = \mathbf{A} \Leftrightarrow \\
 &\Leftrightarrow \begin{bmatrix} \mathbf{A}\gamma_1 & \dots & \mathbf{A}\gamma_n \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix}^t = \mathbf{A} \Leftrightarrow \\
 &\Leftrightarrow \begin{bmatrix} \theta_1\gamma_1 & \dots & \theta_n\gamma_n \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix}^t = \mathbf{A} \Leftrightarrow \\
 &\Leftrightarrow \begin{bmatrix} \theta_1\gamma_1 & \dots & \theta_n\gamma_n \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix}^t = \mathbf{A} \Leftrightarrow \\
 &\Leftrightarrow \mathbf{P}\mathbf{D}(\theta_1, \dots, \theta_n)\mathbf{P}^t = \mathbf{A} \Leftrightarrow \\
 &\Leftrightarrow \sum_{i=1}^n \theta_i \gamma_i \gamma_i^t = \mathbf{A}
 \end{aligned}$$

■

**Corolário 2.1.3.1** Se  $\mathbf{A}$  é uma matriz simétrica então  $\mathbf{P}$  diagonaliza  $\mathbf{A}$ , isto é,

$$\mathbf{P}^t \mathbf{A} \mathbf{P} = \mathbf{D}(\theta_1, \dots, \theta_n).$$

**Dem.** Pelo teorema 2.1.3, sabemos que

$$\mathbf{A} = \mathbf{P}\mathbf{D}(\theta_1, \dots, \theta_n)\mathbf{P}^t$$

logo

$$\begin{aligned}
 \mathbf{A} &= \mathbf{P}\mathbf{D}(\theta_1, \dots, \theta_n)\mathbf{P}^t \Leftrightarrow \mathbf{P}^t \mathbf{A} = \mathbf{P}^t \mathbf{P} \mathbf{D}(\theta_1, \dots, \theta_n) \mathbf{P}^t \Leftrightarrow \\
 &\Leftrightarrow \mathbf{P}^t \mathbf{A} = \mathbf{D}(\theta_1, \dots, \theta_n) \mathbf{P}^t \Leftrightarrow \mathbf{P}^t \mathbf{A} \mathbf{P} = \mathbf{D}(\theta_1, \dots, \theta_n) \mathbf{P}^t \mathbf{P} \Leftrightarrow \\
 &\Leftrightarrow \mathbf{P}^t \mathbf{A} \mathbf{P} = \mathbf{D}(\theta_1, \dots, \theta_n)
 \end{aligned}$$

■

**Teorema 2.1.4** Se  $\mathbf{A}$  é uma matriz com valores próprios  $\theta_1, \dots, \theta_n$ , então:

$$\bullet \quad |\mathbf{A}| = \prod_{i=1}^n \theta_i;$$

$$\bullet \operatorname{tr}(\mathbf{A}) = \sum_{i=1}^n \theta_i,$$

onde  $\operatorname{tr}(\cdot)$  designa o traço de uma matriz, ou seja, o traço de uma matriz é a soma dos seus elementos principais..

**Dem.** Pelo teorema 2.1.1, sabemos que

$$|\mathbf{A}| = |\mathbf{P}^t \mathbf{A} \mathbf{P}|.$$

Assim

$$|\mathbf{A}| = |\mathbf{P}^t \mathbf{A} \mathbf{P}| = |\mathbf{D}(\theta_1, \dots, \theta_n)| = \prod_{i=1}^n \theta_i.$$

Facilmente se prova que

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^n \theta_i.$$

■

### 2.1.3 Matrizes Inversas Generalizadas e Matrizes Inversas de Moore-Penrose

Sempre que uma matriz  $\mathbf{X}$  é singular diz-se que  $\mathbf{X}^{-1}$  não existe, mas existe uma matriz inversa generalizada, que como o seu próprio nome indica é uma generalização da matriz inversa, ver [Schott, 1997], de tal forma que

**Definição 2.1.2 (*Matriz Inversa Generalizada*)** A matriz inversa generalizada de uma matriz  $\mathbf{A}$  do tipo  $m \times n$  é uma matriz  $\mathbf{G}$  do tipo  $n \times m$  e que verifica

$$\mathbf{A} \mathbf{G} \mathbf{A} = \mathbf{A}.$$

Assim, ver [Pollock, 1979], temos

**Teorema 2.1.5** Qualquer que seja a matriz  $\mathbf{A}$  existe uma e uma só matriz  $\mathbf{A}^+$  tal que:

- $\mathbf{A} \mathbf{A}^+ \mathbf{A} = \mathbf{A}$
- $\mathbf{A}^+ \mathbf{A} \mathbf{A}^+ = \mathbf{A}^+$
- $(\mathbf{A} \mathbf{A}^+)^t = \mathbf{A} \mathbf{A}^+$
- $(\mathbf{A}^+ \mathbf{A})^t = \mathbf{A}^+ \mathbf{A}$

A matriz  $\mathbf{A}^+$  denomina-se a inversa de Moore-Penrose de  $\mathbf{A}$ . Se  $\mathbf{A}$  for uma matriz invertível então  $\mathbf{A}^+ = \mathbf{A}^{-1}$ .

A matriz inversa de Moore-Penrose é muito útil em demonstrações envolvendo modelos lineares.

Note-se ainda que sendo  $\mathbf{A}$  uma matriz simétrica de ordem  $k$ , com

$$\text{car}(\mathbf{A}) = l < k, \quad (2.1)$$

onde  $\text{car}(\cdot)$  designa a característica de uma matriz, pode-se sempre ordenar os vectores linha de uma matriz  $\mathbf{P}$  diagonalizadora ortogonal de  $\mathbf{A}$  de forma a ter-se

$$\mathbf{PAP}^t = \mathbf{D}(r_1, \dots, r_l, 0, \dots, 0) \quad (2.2)$$

com  $r_1, \dots, r_l$  os valores próprios não nulos de  $\mathbf{A}$ , tendo-se

$$\mathbf{A} = \mathbf{P}^t \mathbf{D}(r_1, \dots, r_l, 0, \dots, 0) \mathbf{P} \quad (2.3)$$

e

$$\mathbf{A}^+ = \mathbf{P}^t \mathbf{D}(r_1^+, \dots, r_l^+, 0, \dots, 0) \mathbf{P}. \quad (2.4)$$

### 2.1.4 Matrizes de Projecção Ortogonal

Seja  $E$  um espaço vectorial e  $S$  um subespaço vectorial de  $E$ . Consideremos ainda  $\{\mathbf{z}_1, \dots, \mathbf{z}_m\}$  uma base ortonormada de  $E$  e  $\{\mathbf{z}_1, \dots, \mathbf{z}_r\}$  uma base ortonormada de  $S$ , sendo  $r < m$ .

Consideremos ainda  $\mathbf{x} \in E$  tal que

$$\mathbf{x} = \sum_{i=1}^r \alpha_i \mathbf{z}_i + \sum_{i=r+1}^m \alpha_i \mathbf{z}_i. \quad (2.5)$$

**Definição 2.1.3 (Complemento Ortogonal)** *Seja  $S$  um subespaço vectorial de  $E$ . O complemento ortogonal de  $S$ , denotado por  $S^\perp$ , é o conjunto de todos os vectores de  $E$  que são ortogonais a todos os vectores em  $S$ . Assim*

$$S^\perp = \{\mathbf{x} : \mathbf{x} \in E \text{ e } \mathbf{x}^t \mathbf{y} = 0 \text{ para todo o } \mathbf{y} \in S\}.$$

**Teorema 2.1.6** *Se  $S$  é um subespaço vectorial de  $E$  então o seu complemento ortogonal  $S^\perp$  também é um subespaço vectorial de  $E$ .*

A consequência do teorema que se segue, é que se  $S$  é um subespaço vectorial de  $E$  e a sua dimensão for  $r$ , então  $S^\perp$  também é um subespaço vectorial de  $E$  e a dimensão é  $m - r$ .

**Teorema 2.1.7** *Suponhamos que  $\{\mathbf{z}_1, \dots, \mathbf{z}_r, \mathbf{z}_{r+1}, \dots, \mathbf{z}_m\}$  é uma base ortonormalizada de  $E$  e  $\{\mathbf{z}_1, \dots, \mathbf{z}_r\}$  é uma base ortonormalizada para o subespaço vectorial  $S$  de  $E$ . Então  $\{\mathbf{z}_{r+1}, \dots, \mathbf{z}_m\}$  é uma base ortonormalizada para  $S^\perp$ .*

A projecção ortogonal de um vector  $\mathbf{x}$  do tipo  $m \times 1$  para um subespaço vectorial  $S$  de  $E$  pode ser convenientemente expressa na forma matricial.

Consideremos

$$\{\mathbf{z}_1, \dots, \mathbf{z}_r\} \quad (2.6)$$

uma qualquer base ortonormalizada para  $S$ , enquanto

$$\{\mathbf{z}_1, \dots, \mathbf{z}_r, \mathbf{z}_{r+1}, \dots, \mathbf{z}_m\} \quad (2.7)$$

é uma base ortonormalizada para  $E$ . Suponhamos ainda que  $\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_m$  são constantes que satisfazem a seguinte relação

$$\begin{aligned} \mathbf{x} &= \alpha_1 \mathbf{z}_1 + \dots + \alpha_r \mathbf{z}_r + \alpha_{r+1} \mathbf{z}_{r+1} + \dots + \alpha_m \mathbf{z}_m = \\ &= (\alpha_1 \mathbf{z}_1 + \dots + \alpha_r \mathbf{z}_r) + (\alpha_{r+1} \mathbf{z}_{r+1} + \dots + \alpha_m \mathbf{z}_m) = \\ &= \sum_{i=1}^r \alpha_i \mathbf{z}_i + \sum_{i=r+1}^m \alpha_i \mathbf{z}_i = \mathbf{u} + \mathbf{v}, \end{aligned} \quad (2.8)$$

ou seja,  $\mathbf{x}$  pode ser escrito como combinação linear dos elementos da base do subespaço vectorial  $S$  e dos elementos da base do subespaço vectorial  $S^\perp$ , sendo  $\mathbf{x}$  um elemento de  $E$ .

Podemos ainda escrever

$$\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_r \ \alpha_{r+1} \cdots \alpha_m]^t = [\boldsymbol{\alpha}_1^t \ \boldsymbol{\alpha}_2^t]^t \quad \text{e} \quad \mathbf{Z} = [\mathbf{Z}_1 \ \mathbf{Z}_2] \quad (2.9)$$

onde

$$\begin{aligned} \boldsymbol{\alpha}_1 &= [\alpha_1 \cdots \alpha_r]^t ; \quad \boldsymbol{\alpha}_2 = [\alpha_{r+1} \cdots \alpha_m]^t \\ \mathbf{Z}_1 &= [\mathbf{z}_1 \cdots \mathbf{z}_r] ; \quad \mathbf{Z}_2 = [\mathbf{z}_{r+1} \cdots \mathbf{z}_m], \end{aligned} \quad (2.10)$$

sendo

$\mathbf{Z}_1$  — a matriz que tem por colunas os elementos da base de  $S$

$\mathbf{Z}_2$  — a matriz que tem por colunas os elementos da base de  $S^\perp$

$\boldsymbol{\alpha}_1$  — o vector coluna com  $r$  coeficientes, sendo  $r$  a dimensão de  $S$

$\boldsymbol{\alpha}_2$  — o vector coluna com  $m - r$  coeficientes, sendo  $m - r$  a dimensão de  $S^\perp$ .

Então atendendo a (2.8), (2.9) e (2.10), podemos reescrever  $\mathbf{x}$  da seguinte forma

$$\begin{aligned} \mathbf{x} &= \mathbf{Z}\boldsymbol{\alpha} = [\mathbf{Z}_1 \ \mathbf{Z}_2] [\boldsymbol{\alpha}_1^t \ \boldsymbol{\alpha}_2^t]^t = \mathbf{Z}_1 \boldsymbol{\alpha}_1 + \mathbf{Z}_2 \boldsymbol{\alpha}_2 = \\ &= \sum_{i=1}^r \alpha_i \mathbf{z}_i + \sum_{i=r+1}^m \alpha_i \mathbf{z}_i = \mathbf{u} + \mathbf{v} \end{aligned}$$



em que consideramos  $\mathbf{u} = \mathbf{Z}_1 \boldsymbol{\alpha}_1$  e  $\mathbf{v} = \mathbf{Z}_2 \boldsymbol{\alpha}_2$ .

Uma vez que  $\{\mathbf{z}_1, \dots, \mathbf{z}_r, \mathbf{z}_{r+1}, \dots, \mathbf{z}_m\}$  é uma base ortonormada de  $E$ , então  $\mathbf{z}_i^t \mathbf{z}_i = 1$  e  $\mathbf{z}_i^t \mathbf{z}_j = 0$  com  $i \neq j$ ,  $i, j = 1, \dots, m$ , temos que

$$\begin{cases} \mathbf{Z}_1^t \mathbf{Z}_1 = \mathbf{I}_r \\ \mathbf{Z}_2^t \mathbf{Z}_2 = \mathbf{I}_{m-r} \\ \mathbf{Z}_1^t \mathbf{Z}_2 = \mathbf{O}_{r, m-r} \\ \mathbf{Z}_2^t \mathbf{Z}_1 = \mathbf{O}_{m-r, r} \end{cases}$$

e então vem

$$\begin{aligned} \mathbf{Z}_1 \mathbf{Z}_1^t \mathbf{x} &= \mathbf{Z}_1 \mathbf{Z}_1^t (\mathbf{Z} \boldsymbol{\alpha}) = \mathbf{Z}_1 \mathbf{Z}_1^t [\mathbf{Z}_1 \mathbf{Z}_2] [\boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_2^t]^t = \\ &= \begin{bmatrix} \mathbf{Z}_1 \underbrace{\mathbf{Z}_1^t \mathbf{Z}_1} & \mathbf{Z}_1 \underbrace{\mathbf{Z}_1^t \mathbf{Z}_2} \end{bmatrix} [\boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_2^t]^t = [\mathbf{Z}_1 \mathbf{0}^{m-r}] [\boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_2^t]^t = \\ &= \mathbf{Z}_1 \boldsymbol{\alpha}_1 = \mathbf{u}, \end{aligned} \quad (2.11)$$

e assim diremos que o vector  $\mathbf{u}$ , como foi descrito em (2.11), é designado pela projecção ortogonal de  $\mathbf{x}$  em  $S$ , uma vez que  $\mathbf{u} = \mathbf{Z}_1 \mathbf{Z}_1^t \mathbf{x}$ , sendo  $\mathbf{Z}_1$  a matriz que tem por colunas os elementos da base do subespaço vectorial  $S$ .

Analogamente, podemos considerar também

$$\begin{aligned} \mathbf{Z}_2 \mathbf{Z}_2^t \mathbf{x} &= \mathbf{Z}_2 \mathbf{Z}_2^t (\mathbf{Z} \boldsymbol{\alpha}) = \mathbf{Z}_2 \mathbf{Z}_2^t [\mathbf{Z}_1 \mathbf{Z}_2] [\boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_2^t]^t = \\ &= \begin{bmatrix} \mathbf{Z}_2 \underbrace{\mathbf{Z}_2^t \mathbf{Z}_1} & \mathbf{Z}_2 \underbrace{\mathbf{Z}_2^t \mathbf{Z}_2} \end{bmatrix} [\boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_2^t]^t = [\mathbf{0}^r \mathbf{Z}_2] [\boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_2^t]^t = \\ &= \mathbf{Z}_2 \boldsymbol{\alpha}_2 = \mathbf{v}, \end{aligned} \quad (2.12)$$

e diremos que o vector  $\mathbf{v}$ , como foi descrito em (2.12), é designado pela projecção ortogonal de  $\mathbf{x}$  em  $S^\perp$ , uma vez que  $\mathbf{v} = \mathbf{Z}_2 \mathbf{Z}_2^t \mathbf{x}$ , sendo  $\mathbf{Z}_2$  a matriz que tem por colunas os elementos da base do subespaço vectorial  $S^\perp$ .

Após a exposição anterior, podemos considerar o seguinte teorema

**Teorema 2.1.8** *Suponhamos que as colunas da matriz  $\mathbf{Z}_1$  do tipo  $m \times r$  formam uma base ortonormada para o subespaço vectorial  $S$  (de dimensão  $r$ ), que é um subespaço vectorial de  $E$ . Se  $\mathbf{x} \in E$ , então a projecção ortogonal de  $\mathbf{x}$  em  $S$  é dada por  $\mathbf{Z}_1 \mathbf{Z}_1^t \mathbf{x}$ .*

Analogamente, podemos também considerar o seguinte teorema

**Teorema 2.1.9** *Suponhamos que as colunas da matriz  $\mathbf{Z}_2$  do tipo  $m \times (m - r)$  formam uma base ortonormada para o subespaço vectorial  $S^\perp$  (de dimensão  $m - r$ ), que é um subespaço vectorial de  $E$ . Se  $\mathbf{x} \in E$ , então a projecção ortogonal de  $\mathbf{x}$  em  $S^\perp$  é dada por  $\mathbf{Z}_2 \mathbf{Z}_2^t \mathbf{x}$ .*

A matriz  $\mathbf{Z}_1\mathbf{Z}_1^t$  é designada por matriz de projecção ortogonal sobre o subespaço vectorial  $S$  e por vezes pode ser designada por  $\mathbf{P}_S$ . Similarmente,  $\mathbf{Z}_2\mathbf{Z}_2^t$  é a matriz de projecção ortogonal sobre o subespaço vectorial  $S^\perp$ . Assim, é de notar também que

$$\mathbf{Z}\mathbf{Z}^t = [\mathbf{Z}_1 \ \mathbf{Z}_2] \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} = \mathbf{Z}_1\mathbf{Z}_1^t + \mathbf{Z}_2\mathbf{Z}_2^t = \mathbf{I}_m \quad (2.13)$$

é a matriz de projecção ortogonal sobre o espaço vectorial  $E$ .

Uma vez que

$$\mathbf{Z}\mathbf{Z}^t = \mathbf{Z}_1\mathbf{Z}_1^t + \mathbf{Z}_2\mathbf{Z}_2^t, \quad (2.14)$$

e que  $\mathbf{Z}\mathbf{Z}^t = \mathbf{I}_m$ , então temos a equação simples

$$\mathbf{Z}_2\mathbf{Z}_2^t = \mathbf{I}_m - \mathbf{Z}_1\mathbf{Z}_1^t \quad (2.15)$$

relativa às matrizes de projecção sobre o subespaço vectorial e sobre o seu complemento ortogonal,  $S$  e  $S^\perp$ , respectivamente.

Embora um subespaço vectorial não tenha uma base única ortonormada, a matriz de projecção ortogonal formada a partir desta base ortonormada é única. Para provar isto, temos o seguinte teorema.

**Teorema 2.1.10** *Suponhamos que as colunas da matriz  $\mathbf{Z}_1$  do tipo  $m \times r$  formam uma base ortonormada para o subespaço vectorial  $S$  de dimensão  $r$  e que as colunas da matriz  $\mathbf{W}_1$  do tipo  $m \times r$ , também formam uma base ortonormada para o mesmo subespaço vectorial. Então*

$$\mathbf{Z}_1\mathbf{Z}_1^t = \mathbf{W}_1\mathbf{W}_1^t.$$

**Definição 2.1.4 (Matriz de Projecção Ortogonal)** A matriz  $\mathbf{P}_S = \mathbf{Z}_1\mathbf{Z}_1^t$  designa-se por matriz de projecção ortogonal em  $S$ .

**Definição 2.1.5 (Espaço Nulo de uma matriz)** O espaço nulo de uma matriz  $\mathbf{A}$  é o espaço que contém todas as soluções de  $\mathbf{A}\mathbf{x} = \mathbf{0}$ , e representa-se por  $\mathcal{N}(\mathbf{A})$ , ou seja

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}.$$

**Definição 2.1.6 (Espaço Imagem de uma matriz)** O espaço imagem de uma matriz  $\mathbf{A}$  é

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in E\}.$$

Sendo  $\mathbf{P}_S$  a matriz de projecção ortogonal em  $S$ , sabemos que  $\mathcal{R}(\mathbf{P}_S) = S$ , pois

$$\mathcal{R}(\mathbf{P}_S) = \{\mathbf{P}_S \mathbf{x} : \mathbf{x} \in E\} = \{\mathbf{Z}_1 \mathbf{Z}_1^t \mathbf{x} : \mathbf{x} \in E\} = S \quad (2.16)$$

e que  $\mathcal{N}(\mathbf{P}_S) = S^\perp$  porque

$$\mathcal{N}(\mathbf{P}_S) = \{\mathbf{x} : \mathbf{P}_S \mathbf{x} = \mathbf{0}^m\} = \{\mathbf{x} : \mathbf{Z}_1 \mathbf{Z}_1^t \mathbf{x} = \mathbf{0}^m\} = S^\perp, \quad (2.17)$$

ou seja, o espaço imagem de  $\mathbf{P}_S$  é  $S$  e o espaço nulo de  $\mathbf{P}_S$  é  $S^\perp$ .

É fácil ver que a matriz  $\mathbf{P}_S$  é uma matriz simétrica e uma matriz idempotente, pois

$$(\mathbf{P}_S)^t = (\mathbf{Z}_1 \mathbf{Z}_1^t)^t = (\mathbf{Z}_1^t)^t \mathbf{Z}_1^t = \mathbf{Z}_1 \mathbf{Z}_1^t = \mathbf{P}_S \quad (2.18)$$

e

$$\mathbf{P}_S \mathbf{P}_S = \mathbf{Z}_1 \mathbf{Z}_1^t \mathbf{Z}_1 \mathbf{Z}_1^t = \mathbf{Z}_1 \mathbf{I}_r \mathbf{Z}_1^t = \mathbf{Z}_1 \mathbf{Z}_1^t = \mathbf{P}_S. \quad (2.19)$$

De um modo geral, se definirmos a matriz  $\mathbf{X}_1 = [\mathbf{x}_1 \cdots \mathbf{x}_r]$  em que  $(\mathbf{x}_1 \cdots \mathbf{x}_r)$  são colunas linearmente independentes, então as colunas de

$$\mathbf{Z}_1 = \mathbf{X}_1 \mathbf{A} \quad (2.20)$$

formarão uma base ortonormada para  $S$  se  $\mathbf{A}$  é uma qualquer matriz do tipo  $r \times r$  para a qual

$$\mathbf{Z}_1^t \mathbf{Z}_1 = \mathbf{I}_r \Leftrightarrow (\mathbf{X}_1 \mathbf{A})^t (\mathbf{X}_1 \mathbf{A}) = \mathbf{I}_r \Leftrightarrow \mathbf{A}^t \mathbf{X}_1^t \mathbf{X}_1 \mathbf{A} = \mathbf{I}_r. \quad (2.21)$$

Assim, a matriz  $\mathbf{A}$  será uma matriz invertível uma vez que

$$\text{car}(\mathbf{X}_1) = \text{car}(\mathbf{Z}_1) = r \quad (2.22)$$

e por isso  $\mathbf{A}^{-1}$  existe, o que faz com que

$$\begin{aligned} \mathbf{A}^t \mathbf{X}_1^t \mathbf{X}_1 \mathbf{A} &= \mathbf{I}_r \Leftrightarrow (\mathbf{A}^t)^{-1} \mathbf{A}^t \mathbf{X}_1^t \mathbf{X}_1 \mathbf{A} \mathbf{A}^{-1} = (\mathbf{A}^t)^{-1} \mathbf{I}_r \mathbf{A}^{-1} \Leftrightarrow \\ &\Leftrightarrow \mathbf{X}_1^t \mathbf{X}_1 = (\mathbf{A}^t)^{-1} \mathbf{A}^{-1} \Leftrightarrow \mathbf{X}_1^t \mathbf{X}_1 = (\mathbf{A} \mathbf{A}^t)^{-1} \end{aligned} \quad (2.23)$$

ou equivalentemente

$$\mathbf{X}_1^t \mathbf{X}_1 = (\mathbf{A} \mathbf{A}^t)^{-1} \Leftrightarrow (\mathbf{X}_1^t \mathbf{X}_1)^{-1} = \left( (\mathbf{A} \mathbf{A}^t)^{-1} \right)^{-1} \Leftrightarrow (\mathbf{X}_1^t \mathbf{X}_1)^{-1} = \mathbf{A} \mathbf{A}^t. \quad (2.24)$$

Consequentemente, podemos obter a expressão para a matriz de projecção ortogonal  $\mathbf{P}_S$ , sobre o subespaço vectorial  $S$  em termos de  $\mathbf{X}_1$  como sendo

$$\mathbf{P}_S = \mathbf{Z}_1 \mathbf{Z}_1^t = (\mathbf{X}_1 \mathbf{A}) (\mathbf{X}_1 \mathbf{A})^t = \mathbf{X}_1 \mathbf{A} \mathbf{A}^t \mathbf{X}_1^t = \mathbf{X}_1 (\mathbf{X}_1^t \mathbf{X}_1)^{-1} \mathbf{X}_1^t. \quad (2.25)$$

Portanto, já vimos que a matriz de projecção ortogonal em  $S$ , ou seja  $\mathbf{P}_S$ , que é dada por  $\mathbf{Z}_1 \mathbf{Z}_1^t$ , é uma matriz simétrica e idempotente, o que nos leva a afirmar que toda a matriz de projecção ortogonal é simétrica e idempotente.

Vejam agora o teorema que nos diz que toda a matriz simétrica e idempotente  $\mathbf{P}_S$  é uma matriz de projecção ortogonal em  $S$

**Teorema 2.1.11** *Toda a matriz simétrica e idempotente  $\mathbf{P}_S$  é uma matriz de projecção ortogonal em  $S$ .*

O seguinte lema enumera proposições equivalentes sobre matrizes de projecção ortogonal e os subespaços vectoriais associados.

**Lema 2.1.1** *As seguintes proposições são equivalentes:*

1.  $\mathbf{P}$  é uma matriz de projecção ortogonal;
2.  $\mathbf{I}_n - \mathbf{P}$  é uma matriz de projecção ortogonal;
3.  $\mathcal{R}(\mathbf{P}) = \mathcal{N}(\mathbf{I}_n - \mathbf{P})$ ;
4.  $\mathcal{R}(\mathbf{I}_n - \mathbf{P}) = \mathcal{N}(\mathbf{P})$ ;
5.  $\mathcal{R}(\mathbf{P})$  é ortogonal a  $\mathcal{R}(\mathbf{I}_n - \mathbf{P})$ ;
6.  $\mathcal{N}(\mathbf{P})$  é ortogonal a  $\mathcal{N}(\mathbf{I}_n - \mathbf{P})$ .

A soma directa de subespaços vectoriais que se denota por  $\boxplus$ , pode também ser representada pelo espaço imagem de matrizes de projecção ortogonal. Os teoremas que se seguem mostram isso.

**Teorema 2.1.12** *Sejam  $\mathbf{P}_1, \dots, \mathbf{P}_k$  matrizes de projecção ortogonal, tais que  $\mathbf{P}_i \mathbf{P}_j = \mathbf{0}^n$ , com  $i \neq j$ . Então*

1.  $\mathbf{P} = \sum_{i=1}^k \mathbf{P}_i$  é uma matriz de projecção ortogonal;
2.  $\mathcal{R}(\mathbf{P}_i) \cap \mathcal{R}(\mathbf{P}_j) = \{0\}$ , com  $i \neq j$ ;
3.  $\mathcal{R}(\mathbf{P}) = \boxplus_{i=1}^k \mathcal{R}(\mathbf{P}_i)$ .

**Teorema 2.1.13** *Seja  $\mathbf{P}$  uma matriz de projecção ortogonal associado a um subespaço vectorial  $S$ . Suponhamos que  $S$  é uma soma directa de subespaços, isto é,  $S = \boxplus_{i=1}^k S_i$ . Então existem matrizes de projecção ortogonal únicas,  $\mathbf{P}_1, \dots, \mathbf{P}_k$  tais que  $\mathbf{P} = \sum_{i=1}^k \mathbf{P}_i$  e  $\mathbf{P}_i \mathbf{P}_j = \mathbf{0}^n$ , quando  $i \neq j$ .*

É interessante observar que os teoremas 2.1.12 e 2.1.13 complementam-se, na medida em que um é o "contrário" do outro.

Podemos ainda concluir que uma outra forma de obter a matriz de projecção ortogonal sobre um espaço colunas  $\mathcal{R}(\mathbf{X})$  de uma matriz de característica máxima é considerar

$$\mathbf{P}_{\mathcal{R}(\mathbf{X})} = \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t. \quad (2.26)$$

Segue-se a seguinte proposição

**Proposição 2.1.1** *Dada a matriz  $\mathbf{X}$  do tipo  $n \times k$  e tomando-se  $\mathcal{R}(\mathbf{X})$  e  $\mathcal{R}(\mathbf{X}^t)$ , então temos as matrizes de projecção ortogonal sobre  $\mathcal{R}(\mathbf{X})$  e  $\mathcal{R}(\mathbf{X}^t)$  dadas por*

$$\mathbf{P}_{\mathcal{R}(\mathbf{X})} = \mathbf{X} (\mathbf{X}^t \mathbf{X})^+ \mathbf{X}^t \quad e \quad \mathbf{P}_{\mathcal{R}(\mathbf{X}^t)} = (\mathbf{X}^t \mathbf{X})^+ (\mathbf{X}^t \mathbf{X}),$$

*respectivamente.*

### 2.1.5 Matrizes Ortogonais Estandarizadas e Diagonalizadoras Ortogonais

**Definição 2.1.7 (Matriz Ortogonal Estandarizada)** *Seja  $s$  um número inteiro e considerem-se as matrizes ortogonais*

$$\mathbf{P}_s = \begin{bmatrix} \frac{1}{\sqrt{s}} (\mathbf{1}^s)^t \\ \vdots \\ \mathbf{T}_s \end{bmatrix}$$

*em que  $\mathbf{T}_s$  é uma sub-matriz de  $\mathbf{P}_s$  constituída pelas linhas de  $\mathbf{P}_s$  menos a primeira. As linhas de  $\mathbf{T}_s$  são vectores de contrastes mutuamente ortogonais de norma um. Vectores de contrastes são vectores cujas componentes têm soma nula. As matrizes  $\mathbf{P}_s$  assim definidas são chamadas matrizes ortogonais estandarizadas.*

**Definição 2.1.8 (Matriz Diagonalizadora Ortogonal)** *Dada uma matriz simétrica  $\mathbf{M}$  de ordem  $k$ , existe uma matriz  $\mathbf{P}$  ortogonal de ordem  $k$  tal que*

$$\begin{aligned} \mathbf{PMP}^t &= \mathbf{D}(r_1, \dots, r_k) \Leftrightarrow \mathbf{P}^t \mathbf{PMP}^t \mathbf{P} = \mathbf{P}^t \mathbf{D}(r_1, \dots, r_k) \mathbf{P} \Leftrightarrow \\ &\Leftrightarrow \mathbf{M} = \mathbf{P}^t \mathbf{D}(r_1, \dots, r_k) \mathbf{P} \end{aligned}$$

*onde  $\mathbf{D}(r_1, \dots, r_k)$  representa uma matriz diagonal cujos elementos principais,  $r_1, \dots, r_k$ , são os valores próprios da matriz  $\mathbf{M}$ . Neste caso,  $\mathbf{P}$  diz-se a matriz diagonalizadora ortogonal de  $\mathbf{M}$ , e os seus vectores linha serão os vectores próprios ortonormados de  $\mathbf{M}$ , associados aos respectivos valores próprios.*

Caso  $\mathbf{M}$  seja uma matriz definida positiva tem-se  $r_j > 0, j = 1, \dots, k$ , e podemos dizer que está definida a matriz

$$\mathbf{G}_0 = \mathbf{D}(r_1^{-1/2}, \dots, r_k^{-1/2}) \mathbf{P} \quad (2.27)$$

que é solução da equação matricial em  $\mathbf{X}$ ,

$$\mathbf{XMX}^t = \mathbf{I}_k, \quad (2.28)$$

ou seja,

$$\begin{aligned}
 \mathbf{X}\mathbf{M}\mathbf{X}^t &= \mathbf{D} \left( r_1^{-1/2}, \dots, r_k^{-1/2} \right) \mathbf{P}\mathbf{M} \left( \mathbf{D} \left( r_1^{-1/2}, \dots, r_k^{-1/2} \right) \mathbf{P} \right)^t = \\
 &= \mathbf{D} \left( r_1^{-1/2}, \dots, r_k^{-1/2} \right) \mathbf{P}\mathbf{M}\mathbf{P}^t \mathbf{D} \left( r_1^{-1/2}, \dots, r_k^{-1/2} \right) = \\
 &= \mathbf{D} \left( r_1^{-1/2}, \dots, r_k^{-1/2} \right) \mathbf{D} (r_1, \dots, r_k) \mathbf{D} \left( r_1^{-1/2}, \dots, r_k^{-1/2} \right) = \\
 &= \mathbf{D} (1, \dots, 1) = \mathbf{I}_k
 \end{aligned}$$

onde  $\mathbf{I}_k$  é a matriz identidade de ordem  $k$ , e portanto  $\mathbf{G}_0$  é a solução da equação matricial em  $\mathbf{X}$ .

Os vectores linha da matriz  $\mathbf{P}$  formam um conjunto de vectores próprios ortonormados associados aos valores próprios da matriz  $\mathbf{M}$ , sendo possível representar a matriz  $\mathbf{P}$  por

$$\mathbf{P} = [\mathbf{A}_1^t \dots \mathbf{A}_k^t]^t, \quad (2.29)$$

em que  $\mathbf{A}_i^t$  representam os vectores próprios ortonormados associados aos valores próprios da matriz  $\mathbf{M}$ .

Tendo as matrizes simétricas que comutam especial interesse nesta dissertação, apresenta-se o seguinte resultado, ver [Schott, 1997]

**Proposição 2.1.2** *Sejam  $\mathbf{A}_1, \dots, \mathbf{A}_m$  matrizes simétricas de ordem  $k$ . Então existe uma matriz ortogonal  $\mathbf{P}$  tal que para cada  $i$ ,  $\mathbf{P}^t \mathbf{A}_i \mathbf{P} = \mathbf{D}_i$  é uma matriz diagonal cujos elementos principais são os valores próprios de  $\mathbf{A}_i$  se e só se  $\mathbf{A}_i \mathbf{A}_j = \mathbf{A}_j \mathbf{A}_i$  para todos os pares  $(i, j), i, j = 1, \dots, m$ , isto é, se e só se as matrizes  $\mathbf{A}_1, \dots, \mathbf{A}_m$  comutarem.*

### 2.1.6 Produto de Kronecker

O produto de Kronecker de matrizes tem importantes aplicações na Estatística permitindo, na maior parte dos casos, representações mais compactas. Este produto entre matrizes não deve ser confundido com o produto usual, pois este, está definido para quaisquer tipos de matrizes.

Esta operação tem sido largamente estudada por [Graham, 1981], [Steeb, 1991] e [Steeb e Hardy, 2011].

**Definição 2.1.9 (*Produto de Kronecker*)** *Seja  $\mathbf{A} = [a_{i,j}]$  uma matriz do tipo  $m \times n$  e  $\mathbf{B}$  uma matriz do tipo  $p \times q$ , então o produto de Kronecker entre estas duas matrizes,  $\mathbf{A} \otimes \mathbf{B}$ , será dado pela matriz por blocos sendo  $\mathbf{A} \otimes \mathbf{B}$  uma matriz do tipo  $mp \times nq$ , ou seja*

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,n}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}\mathbf{B} & a_{m,2}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{bmatrix},$$

com

$$\mathbf{B} = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,q} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p,1} & b_{p,2} & \cdots & b_{p,q} \end{bmatrix}.$$

Observe-se que o produto de Kronecker de matrizes não goza da propriedade comutativa.

Vejamos agora alguns dos resultados sobre o Produto de Kronecker, que podem ser encontrados em [Schott, 1997].

**Teorema 2.1.14** *Sejam  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  matrizes quaisquer e  $\mathbf{a}$  e  $\mathbf{b}$  vectores quaisquer. Então*

- $\alpha \otimes \mathbf{A} = \mathbf{A} \otimes \alpha = \alpha \mathbf{A}$ , para qualquer escalar  $\alpha$ ;
- $(\alpha \mathbf{A}) \otimes (\beta \mathbf{B}) = \alpha \beta (\mathbf{A} \otimes \mathbf{B})$ , para quaisquer escalares  $\alpha$  e  $\beta$ ;
- $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$ ;
- $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C})$ , se  $\mathbf{A}$  e  $\mathbf{B}$  são do mesmo tipo;
- $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C})$ , se  $\mathbf{B}$  e  $\mathbf{C}$  são do mesmo tipo;
- $\mathbf{a} \mathbf{b}^t = \mathbf{a} \otimes \mathbf{b}^t = \mathbf{b}^t \otimes \mathbf{a}$ .

Tendo em conta as propriedades apresentadas no teorema 2.1.14, facilmente se estabelece que

$$\left( \sum_{i=1}^m a_i \mathbf{U}_i \right) \otimes \left( \sum_{j=1}^n b_j \mathbf{V}_j \right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j (\mathbf{U}_i \otimes \mathbf{V}_j). \quad (2.30)$$

A proposição seguinte fornece um resultado útil de combinação entre o produto usual de matrizes e o produto de Kronecker.

**Teorema 2.1.15** *Sejam  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  e  $\mathbf{D}$  matrizes dos tipos  $m \times h$ ,  $p \times k$ ,  $h \times n$  e  $k \times q$ , respectivamente. Então*

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}.$$

**Corolário 2.1.15.1** *Sejam  $\mathbf{A}$  e  $\mathbf{B}$  matrizes idempotentes quaisquer. Então  $\mathbf{A} \otimes \mathbf{B}$  é uma matriz idempotente.*

**Dem.** Atendendo ao teorema 2.1.15 e admitindo o produto usual definido, com  $\mathbf{A}$  e  $\mathbf{B}$  matrizes idempotentes, temos

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{A} \otimes \mathbf{B}) = \mathbf{A}\mathbf{A} \otimes \mathbf{B}\mathbf{B} = \mathbf{A} \otimes \mathbf{B}.$$

■

**Corolário 2.1.15.2** *Sejam  $\mathbf{A}$  e  $\mathbf{B}$  matrizes ortogonais quaisquer. Então  $\mathbf{A} \otimes \mathbf{B}$  é uma matriz ortogonal.*

**Dem.** Basta observar que sendo  $\mathbf{A}$  e  $\mathbf{B}$  matrizes ortogonais de ordem  $m$  e  $n$ , respectivamente, têm-se

$$\begin{aligned} (\mathbf{A} \otimes \mathbf{B})(\mathbf{A} \otimes \mathbf{B})^t &= (\mathbf{A} \otimes \mathbf{B})(\mathbf{A}^t \otimes \mathbf{B}^t) = \mathbf{A}\mathbf{A}^t \otimes \mathbf{B}\mathbf{B}^t = \\ &= \mathbf{I}_m \otimes \mathbf{I}_n = \mathbf{I}_{mn}. \end{aligned}$$

Analogamente

$$(\mathbf{A} \otimes \mathbf{B})^t(\mathbf{A} \otimes \mathbf{B}) = \mathbf{I}_{mn}.$$

■

**Corolário 2.1.15.3** *Sejam  $\mathbf{A}$  e  $\mathbf{B}$  matrizes de projecção ortogonal. Então  $\mathbf{A} \otimes \mathbf{B}$  é uma matriz de projecção ortogonal.*

**Dem.** Prova-se usando a demonstração de 2.1.15.1 e 2.1.15.2. ■

**Teorema 2.1.16** *Sejam  $\mathbf{A}$  e  $\mathbf{B}$  matrizes quaisquer. Então*

$$(\mathbf{A} \otimes \mathbf{B})^t = \mathbf{A}^t \otimes \mathbf{B}^t.$$

**Dem.** Sabe-se que

$$\begin{aligned} (\mathbf{A} \otimes \mathbf{B})^t &= \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,n}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}\mathbf{B} & a_{m,2}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{bmatrix}^t = \begin{bmatrix} a_{1,1}\mathbf{B}^t & a_{2,1}\mathbf{B}^t & \vdots & a_{m,1}\mathbf{B}^t \\ a_{1,2}\mathbf{B}^t & a_{2,2}\mathbf{B}^t & \vdots & a_{m,2}\mathbf{B}^t \\ \dots & \dots & \ddots & \dots \\ a_{1,n}\mathbf{B}^t & a_{2,n}\mathbf{B}^t & \vdots & a_{m,n}\mathbf{B}^t \end{bmatrix} = \\ &= \begin{bmatrix} a_{1,1} & a_{2,1} & \vdots & a_{m,1} \\ a_{1,2} & a_{2,2} & \vdots & a_{m,2} \\ \dots & \dots & \ddots & \dots \\ a_{1,n} & a_{2,n} & \vdots & a_{m,n} \end{bmatrix} \otimes \mathbf{B}^t = \mathbf{A}^t \otimes \mathbf{B}^t \end{aligned}$$

■



**Corolário 2.1.16.1** *Sejam  $\mathbf{A}$  e  $\mathbf{B}$  matrizes simétricas quaisquer. Então*

$$(\mathbf{A} \otimes \mathbf{B})^t = \mathbf{A} \otimes \mathbf{B},$$

*isto é,  $\mathbf{A} \otimes \mathbf{B}$  é uma matriz simétrica.*

**Dem.** Sendo  $\mathbf{A}$  e  $\mathbf{B}$  matrizes simétricas, tem-se atendendo ao teorema 2.1.16 que

$$(\mathbf{A} \otimes \mathbf{B})^t = \mathbf{A}^t \otimes \mathbf{B}^t = \mathbf{A} \otimes \mathbf{B}.$$

■

Têm-se ainda os seguintes teoremas

**Teorema 2.1.17** *Seja  $\mathbf{A}$  uma matriz quadrada de ordem  $m$  e seja  $\mathbf{B}$  uma matriz quadrada de ordem  $p$ . Então*

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^m \times |\mathbf{B}|^p.$$

**Teorema 2.1.18** *Seja  $\mathbf{A}$  uma matriz quadrada de ordem  $m$  e seja  $\mathbf{B}$  uma matriz quadrada de ordem  $p$ . Então*

$$\text{car}(\mathbf{A} \otimes \mathbf{B}) = \text{car}(\mathbf{A}) \times \text{car}(\mathbf{B}).$$

**Teorema 2.1.19** *Seja  $\mathbf{A}$  uma matriz do tipo  $m \times n$ , e seja  $\mathbf{B}$  uma matriz do tipo  $p \times q$ . Então*

- $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ , se  $m = n, p = q$  e  $\mathbf{A} \otimes \mathbf{B}, \mathbf{A}$  e  $\mathbf{B}$  são matrizes invertíveis.
- $(\mathbf{A} \otimes \mathbf{B})^+ = \mathbf{A}^+ \otimes \mathbf{B}^+$
- $(\mathbf{A} \otimes \mathbf{B})^- = \mathbf{A}^- \otimes \mathbf{B}^-$ , para qualquer inversa generalizada,  $\mathbf{A}^-$  e  $\mathbf{B}^-$ , de  $\mathbf{A}$  e  $\mathbf{B}$ , respectivamente.

Vejamos agora ainda um resultado importante. Consideremos  $\mathbf{D}(\mathbf{a})$  e  $\mathbf{D}(\mathbf{b})$  matrizes diagonais cujos elementos principais são as componentes do vectores  $\mathbf{a}$  e  $\mathbf{b}$ , respectivamente. Então tem-se

$$\mathbf{D}(\mathbf{a}) \otimes \mathbf{D}(\mathbf{b}) = \mathbf{D}(\mathbf{a} \otimes \mathbf{b}). \quad (2.31)$$

Dadas as matrizes simétricas  $\mathbf{M}_1$  e  $\mathbf{M}_2$ , existem matrizes ortogonais  $\mathbf{P}_1$  e  $\mathbf{P}_2$  e matrizes diagonais  $\mathbf{D}(\mathbf{a})$  e  $\mathbf{D}(\mathbf{b})$  tais que

$$\mathbf{P}_1 \mathbf{M}_1 \mathbf{P}_1^t = \mathbf{D}(\mathbf{a}) \quad \text{e} \quad \mathbf{P}_2 \mathbf{M}_2 \mathbf{P}_2^t = \mathbf{D}(\mathbf{b}). \quad (2.32)$$

Os elementos principais de  $\mathbf{D}(\mathbf{a})$  e  $\mathbf{D}(\mathbf{b})$  são os valores próprios de  $\mathbf{M}_1$  e  $\mathbf{M}_2$ , respectivamente, correspondendo-lhes como vectores próprios, os vectores linha de  $\mathbf{P}_1$  e  $\mathbf{P}_2$ . Assim, tem-se

$$\begin{aligned} (\mathbf{P}_1 \otimes \mathbf{P}_2) (\mathbf{M}_1 \otimes \mathbf{M}_2) (\mathbf{P}_1 \otimes \mathbf{P}_2)^t &= (\mathbf{P}_1 \mathbf{M}_1 \mathbf{P}_1^t) \otimes (\mathbf{P}_2 \mathbf{M}_2 \mathbf{P}_2^t) = \\ &= \mathbf{D}(\mathbf{a}) \otimes \mathbf{D}(\mathbf{b}) = \mathbf{D}(\mathbf{a} \otimes \mathbf{b}). \end{aligned}$$

pelo que os valores próprios de  $\mathbf{M}_1 \otimes \mathbf{M}_2$  serão os produtos dos valores próprios  $\mathbf{M}_1$  e  $\mathbf{M}_2$ , e os vectores próprios são os vectores linha de  $\mathbf{P}_1 \otimes \mathbf{P}_2$ .

Vejamos o seguinte teorema que também pode ser encontrado em [Schott, 1997], bem como a sua demonstração.

**Teorema 2.1.20** *Sejam  $\lambda_1, \dots, \lambda_m$  os valores próprios de uma matriz  $\mathbf{A}$  quadrada de ordem  $m$  e sejam  $\theta_1, \dots, \theta_p$  os valores próprios de uma matriz  $\mathbf{B}$  quadrada de ordem  $p$ . Então o conjunto dos  $mp$  valores próprios de  $\mathbf{A} \otimes \mathbf{B}$  é dado por*

$$\{\lambda_i \theta_j : i = 1, \dots, m; j = 1, \dots, p\}.$$

### 2.1.7 Operador vec

Às vezes é conveniente reorganizar os elementos de uma matriz  $\mathbf{A} = [a_{ij}]$  do tipo  $m \times n$ , em um vector coluna do tipo  $mn \times 1$ . Uma forma de o fazer, é empilhar sucessivamente as colunas da matriz  $\mathbf{A}$ , e obtém-se o operador  $\text{vec}(\mathbf{A})$ . Assim, sendo

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1(n-1)} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2(n-1)} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{(m-1)1} & a_{(m-1)2} & \cdots & a_{(m-1)(n-1)} & a_{(m-1)n} \\ a_{m1} & a_{m2} & \cdots & a_{m(n-1)} & a_{mn} \end{bmatrix} = \\ &= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_{n-1} & \mathbf{a}_n \end{bmatrix} \end{aligned}$$

em que  $\mathbf{a}_1, \dots, \mathbf{a}_n$  são os vectores coluna de  $\mathbf{A}$ , vem

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_{n-1} \\ \mathbf{a}_n \end{bmatrix}. \quad (2.33)$$

Note-se que, por definição, o elemento na posição  $(i, j)$  da matriz  $\mathbf{A}$  é o  $(j-1)m + i$ -ésimo elemento de  $\text{vec}(\mathbf{A})$ , ver [Harville, 1998].

Vejam algumas propriedades deste operador

- $\text{vec}(\mathbf{a}^t) = \text{vec}(\mathbf{a}) = \mathbf{a}$ , sendo  $\mathbf{a}$  um vector coluna qualquer;
- $\text{vec}(\mathbf{b}\mathbf{a}^t) = \mathbf{a} \otimes \mathbf{b}$ , sendo  $\mathbf{a}$  um vector coluna  $n \times 1$  e  $\mathbf{b}$  um vector coluna  $m \times 1$ ;
- $\text{vec}(c\mathbf{A}) = c\text{vec}(\mathbf{A})$ , sendo  $c$  um escalar e  $\mathbf{A}$  uma matriz qualquer;
- $\text{vec}(\mathbf{A} + \mathbf{B}) = \text{vec}(\mathbf{A}) + \text{vec}(\mathbf{B})$ , sendo  $\mathbf{A}$  e  $\mathbf{B}$  matrizes do mesmo tipo;
- $\text{vec}\left(\sum_{i=1}^n c_i \mathbf{A}_i\right) = \sum_{i=1}^n c_i \text{vec}(\mathbf{A}_i)$ , sendo  $c_i$  um escalar e  $\mathbf{A}_i$  uma matriz qualquer, com  $i = 1, \dots, n$ .

Seja  $\mathbf{A}$  uma matriz do tipo  $m \times n$  e  $\mathbf{B}$  uma matriz do tipo  $n \times p$  em que  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p$  representam as  $p$  colunas da matriz  $\mathbf{B}$ . Então

$$\text{vec}(\mathbf{AB}) = \begin{bmatrix} \mathbf{A}\mathbf{b}_1 \\ \mathbf{A}\mathbf{b}_2 \\ \vdots \\ \mathbf{A}\mathbf{b}_p \end{bmatrix} \quad (2.34)$$

que como é evidente a  $j$ -ésima coluna da matriz  $\mathbf{AB}$  é  $\mathbf{A}\mathbf{b}_j$ . Note-se que podemos também considerar

$$\text{vec}(\mathbf{AB}) = D(\mathbf{A}, \mathbf{A}, \dots, \mathbf{A}) \text{vec}(\mathbf{B}) \quad (2.35)$$

ou equivalentemente

$$\text{vec}(\mathbf{AB}) = (\mathbf{I}_p \otimes \mathbf{A}) \text{vec}(\mathbf{B}). \quad (2.36)$$

Generalizando este resultado, temos o seguinte teorema

**Teorema 2.1.21** *Para quaisquer matriz  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  do tipo  $m \times n$ ,  $n \times p$  e  $p \times q$ , respectivamente,*

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^t \otimes \mathbf{A}) \text{vec}(\mathbf{B})$$

Assim, atendendo ao teorema 2.1.21, podemos também expressar o  $\text{vec}$  do produto  $\mathbf{AB}$  das duas formas seguintes

- $\text{vec}(\mathbf{AB}) = (\mathbf{B}^t \otimes \mathbf{I}_m) \text{vec}(\mathbf{A})$ ;
- $\text{vec}(\mathbf{AB}) = (\mathbf{B}^t \otimes \mathbf{A}) \text{vec}(\mathbf{I}_n)$ .

**Teorema 2.1.22** Para quaisquer matrizes  $\mathbf{A}$  e  $\mathbf{B}$  do tipo  $m \times n$ , temos

$$\text{tr}(\mathbf{A}^t \mathbf{B}) = (\text{vec}(\mathbf{A}))^t (\text{vec}(\mathbf{B}))$$

**Dem.** Consideremos  $\mathbf{a}_j$  e  $\mathbf{b}_j$  as colunas das matrizes  $\mathbf{A}$  e  $\mathbf{B}$ , respectivamente. Observe-se que o  $j$ -ésimo elemento da diagonal de  $\mathbf{A}^t \mathbf{B}$  é igual a  $\mathbf{a}_j^t \mathbf{b}_j$  então temos

$$\begin{aligned} \text{tr}(\mathbf{A}^t \mathbf{B}) &= \sum_{j=1}^n \mathbf{a}_j^t \mathbf{b}_j = \begin{bmatrix} \mathbf{a}_1^t & \mathbf{a}_2^t & \cdots & \mathbf{a}_n^t \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} = \\ &= \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}^t \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} = (\text{vec}(\mathbf{A}))^t (\text{vec}(\mathbf{B})). \end{aligned}$$

■

Note-se, que pelo teorema 2.1.22, podemos dizer que o produto interno usual de quaisquer duas matrizes do tipo  $m \times n$  é igual ao produto interno usual dos seus  $\text{vec}'s$ . Observe-se também que

$$\text{tr}(\mathbf{A}^t \mathbf{B}) = \text{tr}(\mathbf{B}^t \mathbf{A}) = \text{tr}(\mathbf{BA}^t) = \text{tr}(\mathbf{AB}^t). \quad (2.37)$$

Podemos generalizar o teorema 2.1.22

**Teorema 2.1.23** Para quaisquer matriz  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  e  $\mathbf{D}$  do tipo  $m \times n$ ,  $m \times p$ ,  $p \times q$  e  $n \times q$ , respectivamente,

$$\text{tr}(\mathbf{A}^t \mathbf{BCD}^t) = (\text{vec}(\mathbf{A}))^t (\mathbf{D} \otimes \mathbf{B}) (\text{vec}(\mathbf{C}))$$

**Dem.** Usando os teoremas 2.1.21 e 2.1.22

$$\begin{aligned} \text{tr}(\mathbf{A}^t \mathbf{BCD}^t) &= \text{tr}(\mathbf{A}^t (\mathbf{BCD}^t)) = (\text{vec}(\mathbf{A}))^t (\text{vec}(\mathbf{BCD}^t)) = \\ &= (\text{vec}(\mathbf{A}))^t \left( (\mathbf{D}^t)^t \otimes \mathbf{B} \right) \text{vec}(\mathbf{C}) = \\ &= (\text{vec}(\mathbf{A}))^t (\mathbf{D} \otimes \mathbf{B}) (\text{vec}(\mathbf{C})) \end{aligned}$$

■

Note-se que

$$\text{tr}(\mathbf{A}^t \mathbf{BCD}^t) = \text{tr}(\mathbf{D}^t \mathbf{A}^t \mathbf{BC}) = \text{tr}(\mathbf{CD}^t \mathbf{A}^t \mathbf{B}) = \text{tr}(\mathbf{BCD}^t \mathbf{A}^t)$$

e

$$\begin{aligned} \text{tr}(\mathbf{A}^t \mathbf{BCD}^t) &= \text{tr}[(\mathbf{A}^t \mathbf{BCD}^t)^t] = \text{tr}(\mathbf{DC}^t \mathbf{B}^t \mathbf{A}) = \text{tr}(\mathbf{ADC}^t \mathbf{B}^t) = \\ &= \text{tr}(\mathbf{B}^t \mathbf{ADC}^t) = \text{tr}(\mathbf{C}^t \mathbf{B}^t \mathbf{AD}). \end{aligned}$$

## 2.2 Álgebras de Jordan Comutativas

As Álgebras de Jordan Comutativas foram introduzidas em [Jordan *et al.*, 1934], aquando de uma reformulação da Mecânica Quântica. Posteriormente foram redescobertas por Seely ([Seely, 1970], [Seely, 1971], [Seely, 1972], [Seely, 1977]), que as utilizou no âmbito da Inferência Estatística, e iniciou-se assim uma linha de investigação muito interessante, sendo mais tarde usadas, por exemplo, por Zmyślony [Zmyślony, 1978], Drygas [Drygas *et al.*, 1992], Vanaleuween [Vanaleuween *et al.*, 1998], [Vanaleuween *et al.*, 1999], Fonseca [Fonseca *et al.*, 2006], [Fonseca *et al.*, 2007], [Fonseca *et al.*, 2008] e Jesus [Jesus *et al.*, 2009], [Jesus *et al.*, 2009a].

Em [Seely, 1970], Seely designou as Álgebras de Jordan Comutativas como espaços vectoriais quadráticos, o que também é feito por Rao em [Rao e Kleffe, 1988], mas nesta dissertação iremos designá-las como Álgebras de Jordan Comutativas.

**Definição 2.2.1 (Álgebra de Jordan Comutativa)** *Um espaço linear de matrizes reais simétricas,  $\mathcal{A}$ , diz-se uma Álgebra de Jordan Comutativa, se, além de ser uma Álgebra Jordan, dadas quaisquer matrizes  $\mathbf{A}$  e  $\mathbf{B}$  de  $\mathcal{A}$ , temos que  $\mathbf{AB} = \mathbf{BA}$ .*

De um ponto vista menos formal, diremos que as Álgebras de Jordan Comutativas de matrizes simétricas são espaços lineares constituídos por matrizes simétricas que comutam e que contém os quadrados das suas matrizes.

Todas as Álgebras de Jordan Comutativas de matrizes simétricas têm uma só base, a base principal, constituída por matrizes de projecção ortogonal, mutuamente ortogonais, ver [Seely, 1971].

Pode-se ainda mostrar que toda a família de matrizes de projecção ortogonal, mutuamente ortogonais constitui a base principal da Álgebra de Jordan Comutativa de matrizes simétricas, constituída pelas suas combinações lineares.

Assim, dada uma matriz ortogonal

$$\mathbf{P} = [\mathbf{A}_1^t \dots \mathbf{A}_m^t]^t, \quad (2.38)$$

as matrizes

$$\mathbf{Q}_j = \mathbf{A}_j^t \mathbf{A}_j, \quad j = 1, \dots, m \quad (2.39)$$

são matrizes de projecção ortogonal e mutuamente ortogonais, constituindo a base principal de uma Álgebra de Jordan Comutativa de matrizes simétricas,  $\mathcal{A}(\mathbf{P})$ .

Designando por  $\mathcal{A}(\mathbf{P})$  a Álgebra de Jordan Comutativa de matrizes simétricas, representaremos a base principal da Álgebra de Jordan Comutativa de matrizes simétricas por

$$\underline{\mathbf{Q}} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_m\} = bp(\mathcal{A}(\mathbf{P})). \quad (2.40)$$

Inversamente, se considerarmos

$$\mathbf{Q}_j = \mathbf{A}_j^t \mathbf{A}_j, \quad j = 1, \dots, m, \quad (2.41)$$

então as matrizes  $\mathbf{Q}_1, \dots, \mathbf{Q}_m$  constituem a base principal de uma Álgebra de Jordan Comutativa de matrizes simétricas,  $\mathcal{A}(\mathbf{P})$ , sendo

$$\mathbf{P} = [\mathbf{A}_1^t \dots \mathbf{A}_m^t]^t \quad (2.42)$$

a matriz ortogonal associada à álgebra  $\mathcal{A}(\mathbf{P})$ .

Assim, temos

$$\mathbf{A}_j^t \mathbf{A}_j = \mathbf{Q}_j \quad \text{e} \quad \mathbf{A}_j \mathbf{A}_j^t = \mathbf{I}_{g_j}, \quad \text{com } j = 1, \dots, m, \quad (2.43)$$

sendo  $g_j = \text{car}(\mathbf{Q}_j)$ .

Além disso, dado que  $\mathbf{Q}_1, \dots, \mathbf{Q}_m$  são mutuamente ortogonais, teremos

$$\mathbf{Q}_j \mathbf{Q}_i^t = \mathbf{O}_{g_j, g_i} \quad (2.44)$$

onde  $\mathbf{O}_{g_j, g_i}$  representa a matriz nula do tipo  $g_j \times g_i$ .

Observe-se ainda, que os espaços imagem das matrizes da base principal serão também mutuamente ortogonais, ou seja

$$\mathcal{R}(\mathbf{Q}_j) = \{\mathbf{v} : \mathbf{v} = \mathbf{Q}_j \mathbf{u}\}, \quad \text{com } j = 1, \dots, m, \quad (2.45)$$

serão também mutuamente ortogonais, pois

$$(\mathbf{Q}_i \mathbf{u}_i)^t (\mathbf{Q}_j \mathbf{u}_j) = \mathbf{u}_i^t \mathbf{Q}_i^t \mathbf{Q}_j^t \mathbf{u}_j \underset{\mathbf{Q}_1, \dots, \mathbf{Q}_m \text{ são mutuamente ortogonais}}{=} 0$$

Em relação à ordem da matriz  $\mathbf{P}$  com que iremos trabalhar é ordem  $n$ , caso nada seja dito em contrário.

**Definição 2.2.2 (Álgebra de Jordan Comutativa Completa)** Sejam  $g_1, \dots, g_m$  as características das matrizes da base principal  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$  de uma Álgebra de Jordan Comutativa  $\mathcal{A}(\mathbf{P})$ , se

$$\sum_{j=1}^m \mathbf{Q}_j = \mathbf{I}_n \in \mathcal{A}(\mathbf{P})$$

com  $\mathbf{I}_n$  a matriz identidade de ordem  $n$ , diz-se que  $\mathcal{A}(\mathbf{P})$  é completa, o que significa que

$$\sum_{j=1}^m g_j = n,$$

caso contrário, tem-se  $\sum_{j=1}^m g_j < n$ .

Observe-se que podemos dizer que uma Álgebra de Jordan Comutativa de matrizes simétricas é completa se contém as matrizes inversas, pois se  $\mathcal{A}(\mathbf{P})$  contém as matrizes inversas devemos ter

$$\sum_{j=1}^m \mathbf{Q}_j = \mathbf{I}_n \quad (2.46)$$

uma vez que

$$\sum_{j=1}^m g_j = n. \quad (2.47)$$

No caso em que a álgebra não é completa, podemos juntar às matrizes da base principal, a matriz dada por

$$\mathbf{Q}_{m+1} = \mathbf{I}_n - \sum_{j=1}^m \mathbf{Q}_j. \quad (2.48)$$

Assim, neste caso, diz-se que a Álgebra de Jordan Comutativa,  $\overline{\mathcal{A}(\mathbf{P})}$ , com base principal  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m, \mathbf{Q}_{m+1}\}$  é a **Álgebra Completada de  $\mathcal{A}(\mathbf{P})$** .

**Definição 2.2.3 (Álgebra de Jordan Comutativa Regular)** Se

$$\frac{1}{n} \mathbf{J}_n = \frac{1}{n} \mathbf{1}^n (\mathbf{1}^n)^t \in \mathcal{A}(\mathbf{P}),$$

com  $\mathbf{J}_n$  a matriz de ordem  $n$  cujos elementos são todos iguais a 1 e  $\mathbf{1}^n$  o vector coluna de ordem  $n$  cujos elementos também são todos iguais a 1, diz-se que  $\mathcal{A}(\mathbf{P})$  é regular.

Assim, seja  $\mathbf{M}$  uma matriz de projecção ortogonal, portanto,  $\mathbf{M}$  é uma matriz simétrica e idempotente, pertencente a  $\mathcal{A}(\mathbf{P})$ , logo ter-se-á

$$\mathbf{M} = \sum_{j=1}^m a_j \mathbf{Q}_j, \quad (2.49)$$

ou seja,  $\mathbf{M}$  será uma matriz que é combinação linear das matrizes da base principal da álgebra.

Como  $\mathbf{M}$  é idempotente e as matrizes  $\mathbf{Q}_1, \dots, \mathbf{Q}_m$  são mutuamente ortogonais tem-se

$$\sum_{j=1}^m a_j \mathbf{Q}_j = \sum_{j=1}^m a_j^2 \mathbf{Q}_j, \quad (2.50)$$

vindo

$$\sum_{j=1}^m a_j \mathbf{Q}_j = \sum_{j=1}^m a_j^2 \mathbf{Q}_j \Leftrightarrow a_j = a_j^2,$$

pelo que  $a_j = 0$  ou  $a_j = 1$  com  $j = 1, \dots, m$ , e toda a matriz de projecção ortogonal pertencente a  $\mathcal{A}(\mathbf{P})$  é soma de matrizes da base principal. Logo se uma matriz de projecção ortogonal com característica 1 pertencer a uma Álgebra de Jordan Comutativa, pertence à respectiva base principal. Em particular, numa Álgebra de Jordan Comutativa Regular tem-se

$$\mathbf{Q}_1 = \frac{1}{n} \mathbf{J}_n. \quad (2.51)$$

Consideremos agora  $\Upsilon = \{j : a_j \neq 0\} = \{j : a_j = 1\}$ , e atendendo a (2.49) podemos reescrever  $\mathbf{M}$  da seguinte forma

$$\mathbf{M} = \sum_{j \in \Upsilon} \mathbf{Q}_j \quad (2.52)$$

pelo que concluímos que as matrizes de projecção ortogonal pertencentes a  $\mathcal{A}(\mathbf{P})$  serão somas de todas as matrizes da base principal,  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$ , ou então serão a soma de um subconjunto do conjunto das matrizes dessa base, dependendo se existe algum  $a_j = 0$ .

Assim sendo  $\mathbf{M}$  como referido em (2.49) e seja  $\nabla_j$  o espaço imagem da matriz da base principal  $\mathbf{Q}_j$ , temos

$$\mathbf{M} = \sum_{j=1}^m a_j \mathbf{Q}_j \quad \text{e} \quad \nabla_j = \mathcal{R}(\mathbf{Q}_j), \text{ com } j = 1, \dots, m.$$



e representando por  $\boxplus$  a soma directa ortogonal de subespaços vectoriais, tem-se

$$\mathcal{R}(\mathbf{M}) = \mathcal{R}\left(\sum_{j=1}^m a_j \mathbf{Q}_j\right) = \sum_{j=1}^m a_j \mathcal{R}(\mathbf{Q}_j) = \sum_{j \in \Upsilon} \mathcal{R}(\mathbf{Q}_j) = \boxplus_{j \in \Upsilon} \nabla_j,$$

uma vez que os subespaços  $\nabla_j$ , com  $j = 1, \dots, m$  são ortogonais entre si, como vimos em (2.45), logo ter-se-á

$$\text{car}(\mathbf{M}) = \sum_{j \in \Upsilon} g_j \quad (2.53)$$

pois

$$\begin{aligned} \text{car}(\mathbf{M}) &= \text{car}\left(\sum_{j=1}^m a_j \mathbf{Q}_j\right) = \dim(\mathcal{R}(\mathbf{M})) = \sum_{j \in \Upsilon} \dim(\nabla_j) = \\ &= \sum_{j \in \Upsilon} \text{car}(\mathbf{Q}_j) = \sum_{j \in \Upsilon} g_j \end{aligned}$$

com  $g_j = \text{car}(\mathbf{Q}_j) = \dim(\nabla_j)$  com  $j = 1, \dots, m$ .

**Proposição 2.2.1** *Seja*

$$\mathbf{M} = \sum_{j=1}^m a_j \mathbf{Q}_j$$

com  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$  a base principal de uma Álgebra de Jordan Comutativa completa,  $\mathcal{A}(\mathbf{P})$ , então  $a_1, \dots, a_m$  serão os valores próprios de  $\mathbf{M}$  com multiplicidades  $g_1 = \text{car}(\mathbf{Q}_1), \dots, g_m = \text{car}(\mathbf{Q}_m)$ , respectivamente, caso os  $a_1, \dots, a_m$  sejam distintos. Se houver grupos de  $a_j$  idênticos somam-se as respectivas multiplicidades.<sup>1</sup>

**Dem.** Sendo as matrizes  $\mathbf{Q}_i$ , com  $i = 1, \dots, m$ , matrizes de projecção ortogonal de ordem  $n$ , logo simétricas e idempotentes, os seus valores próprios são iguais a 1 ou iguais a 0<sup>2</sup>. Os valores próprios iguais a 1 têm multiplicidade  $g_i$  e os valores próprios iguais a 0 têm multiplicidade  $n - g_i$ , com  $i = 1, \dots, m$ , tendo-se

$$\sum_{j=1}^m g_j = n$$

---

<sup>1</sup>Uma vez que  $\mathbf{M}$  é uma matriz de projecção ortogonal, e por isso é simétrica, então, sabe-se que os seus valores próprios têm idênticas multiplicidades (algébrica e geométrica).

<sup>2</sup>Todos os valores próprios de uma matriz simétrica e idempotente são iguais a 0 ou iguais a 1.

porque a Álgebra de Jordan Comutativa é completa, logo a soma de todas as características das matrizes que constituem a base principal da álgebra será igual à ordem da matriz, bem como

$$\sum_{j=1}^m \mathbf{Q}_j = \mathbf{I}_n.$$

Dados os  $g_i$  vectores próprios de  $\mathbf{Q}_i$  associados ao valor próprio 1, os mesmos constituem uma base ortonormada para  $\nabla_i = \mathcal{R}(\mathbf{Q}_i)$ , com  $i = 1, \dots, m$ , sendo estes subespaços mutuamente ortogonais. Juntando as bases dos subespaços  $\nabla_i$ , com  $i = 1, \dots, m$ , obtém-se uma base

$$\{\mathbf{v}_{1,1}, \dots, \mathbf{v}_{g_1,1}, \mathbf{v}_{1,2}, \dots, \mathbf{v}_{g_2,2}, \dots, \mathbf{v}_{1,m}, \dots, \mathbf{v}_{g_m,m}\}$$

para  $\mathbb{R}^n$ . Nesta base de  $\mathbb{R}^n$ , os primeiros  $g_1$  vectores constituem uma base para  $\nabla_1$ , e assim sucessivamente, até que os últimos  $g_m$  vectores constituem uma base para  $\nabla_m$ . Ora, sendo

$$\mathbf{M} = \sum_{j=1}^m a_j \mathbf{Q}_j,$$

então a matriz cujos vectores linha são os vectores desta base de  $\mathbb{R}^n$  é a matriz diagonalizadora ortogonal  $\mathbf{P}$ , tendo-se

$$\mathbf{PMP}^t = \mathbf{D}(a_1, \dots, a_1, \dots, a_m, \dots, a_m).$$

Logo

$$\underbrace{a_1, \dots, a_1}_{g_1 \text{ vezes}}, \underbrace{a_2, \dots, a_2}_{g_2 \text{ vezes}}, \dots, \underbrace{a_m, \dots, a_m}_{g_m \text{ vezes}},$$

são os valores próprios de  $\mathbf{M}$  com multiplicidades  $g_1, \dots, g_m$ . Os vectores próprios associados a estes valores próprios são os vectores linha

$$\underbrace{v_{1,1}, \dots, v_{g_1,1}}_{g_1 \text{ vezes}}, \underbrace{v_{1,2}, \dots, v_{g_2,2}}_{g_2 \text{ vezes}}, \dots, \underbrace{v_{1,m}, \dots, v_{g_m,m}}_{g_m \text{ vezes}}$$

de  $\mathbf{P}$ , o que estabelece a tese. ■

Se  $\mathbf{Q}$  for uma matriz de projecção ortogonal, portanto, uma matriz simétrica e idempotente, tem-se  $\mathbf{Q}^+ = \mathbf{Q}$ .<sup>3</sup>

Além disso, se

$$\mathbf{M} = \sum_{j=1}^m a_j \mathbf{Q}_j \in \mathcal{A}(\mathbf{P}) \quad (2.54)$$

---

<sup>3</sup>Se  $\mathbf{Q}$  for uma matriz simétrica e idempotente, significa que a Inversa de Moore Penrose coincide com a própria matriz.

com  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$  a base principal de  $\mathcal{A}(\mathbf{P})$  ter-se-á, uma vez que  $\mathbf{Q}_1, \dots, \mathbf{Q}_m$  são idempotentes e mutuamente ortogonais

$$\mathbf{M}^+ = \sum_{j=1}^m a_j^+ \mathbf{Q}_j \in \mathcal{A}(\mathbf{P}) \quad (2.55)$$

com  $a_j^+ = 0$  se  $a_j = 0$  e  $a_j^+ = a_j^{-1}$  se  $a_j \neq 0$ , com  $j = 1, \dots, m$ , ou seja,  $\mathbf{M}^+$  é a inversa de Moore-Penrose de  $\mathbf{M}$ .

Segue-se a seguinte proposição:

**Proposição 2.2.2** *A Álgebra de Jordan Comutativa  $\mathcal{A}(\mathbf{P})$  contém matrizes invertíveis se e só se for completa.*

**Dem.** Seja  $\mathbf{M} \in \mathcal{A}(\mathbf{P})$  uma matriz invertível, então  $\mathbf{M} = \sum_{j \in \Upsilon} a_j \mathbf{Q}_j$  com  $\Upsilon = \{j : a_j \neq 0\}$ , logo atendendo a (2.53) e sendo  $\mathbf{M}$  uma matriz invertível, temos

$$n = \text{car}(\mathbf{M}) = \sum_{j \in \Upsilon} g_j = \sum_{j \in \Upsilon} \text{car}(\mathbf{Q}_j),$$

vindo  $\Upsilon = \{j : a_j \neq 0\} = \{1, \dots, m\}$  e  $\sum_{j=1}^m \mathbf{Q}_j = \mathbf{I}_n$ , logo  $\mathcal{A}(\mathbf{P})$  é uma álgebra completa.

Contrariamente, sendo  $\mathcal{A}(\mathbf{P})$  uma álgebra completa, então tem-se que

$$\mathbf{I}_n = \sum_{j=1}^m \mathbf{Q}_j \in \mathcal{A}(\mathbf{P})$$

e  $\mathbf{I}_n$  é uma matriz invertível, pois  $|\mathbf{I}_n| = 1 \neq 0$ , logo  $\mathcal{A}(\mathbf{P})$  contém matrizes invertíveis. ■

Segue-se a seguinte proposição:

**Proposição 2.2.3** *Sendo  $\mathcal{A}(\mathbf{P})$  uma Álgebra de Jordan Comutativa completa,*

$$\mathbf{M} = \sum_{j=1}^m a_j \mathbf{Q}_j \in \mathcal{A}(\mathbf{P})$$

*diremos que  $\mathbf{M}$  é uma matriz invertível se e só se  $a_j \neq 0$ , com  $j = 1, \dots, m$ .*

**Dem.** Através da proposição 2.2.1 sabemos que  $a_1, \dots, a_m$  são os valores próprios da matriz  $\mathbf{M}$  e como a matriz  $\mathbf{M}$  é invertível, todos os seus valores próprios são não nulos.

Também facilmente se vê que, caso  $\mathbf{M}$  seja invertível então  $|\mathbf{M}| \neq 0$ . Além disso

$$|\mathbf{M}| = \prod_{j=1}^m a_j^{g_j},$$

pois repare-se que

$$\mathbf{PMP}^t = \mathbf{D}(a_1\mathbf{I}_{g_1}, \dots, a_m\mathbf{I}_{g_m}) \Leftrightarrow \mathbf{M} = \mathbf{P}^t\mathbf{D}(a_1\mathbf{I}_{g_1}, \dots, a_m\mathbf{I}_{g_m})\mathbf{P}.$$

Logo

$$\begin{aligned} |\mathbf{M}| &= \left| \sum_{j=1}^m a_j \mathbf{Q}_j \right| = |\mathbf{P}^t \mathbf{D}(a_1\mathbf{I}_{g_1}, \dots, a_m\mathbf{I}_{g_m}) \mathbf{P}| = \\ &= |\mathbf{P}^t| |\mathbf{D}(a_1\mathbf{I}_{g_1}, \dots, a_m\mathbf{I}_{g_m})| |\mathbf{P}| = \prod_{j=1}^m a_j^{g_j} \end{aligned}$$

Como  $|\mathbf{M}| \neq 0$ , então  $a_j \neq 0$ , com  $j = 1, \dots, m$ . ■

Sendo  $\mathbf{Q}_1, \dots, \mathbf{Q}_m$  matrizes de projecção ortogonal, logo simétricas e idempotentes, tem-se

$$\mathbf{Q}_j = \mathbf{A}_j^t \mathbf{A}_j \quad (2.56)$$

com  $\mathbf{A}_j \mathbf{A}_j^t = \mathbf{I}_{g_j}$ , sendo os vectores linha de  $\mathbf{A}_j$  uma base ortonormada para  $\nabla_j = \mathcal{R}(\mathbf{Q}_j)$ , com  $j = 1, \dots, m$ . Portanto, se  $\mathcal{A}(\mathbf{P})$  for uma álgebra completa, ou seja, se

$$\mathbf{I}_n = \sum_{j=1}^m \mathbf{Q}_j, \quad (2.57)$$

os vectores linha da matriz  $\mathbf{P} = [\mathbf{A}_1^t \dots \mathbf{A}_m^t]^t$  constituirão uma base ortonormada de  $\mathbb{R}^n$ , visto ter-se

$$\mathbf{I}_n = \sum_{j=1}^m \mathbf{Q}_j = \sum_{j=1}^m \mathbf{A}_j^t \mathbf{A}_j. \quad (2.58)$$

Então por definição,  $\mathbf{P}$  será uma matriz ortogonal que está associada à álgebra  $\mathcal{A}(\mathbf{P})$ .

Inversamente, dada uma matriz ortogonal  $\mathbf{P}$  e uma partição de  $\bar{n} = \{1, \dots, n\}$  em  $m$  conjuntos  $\Upsilon_j$  com  $j = 1, \dots, m$ , disjuntos dois a dois, ter-se-á

$$\mathbf{I}_n = \sum_{j=1}^m \mathbf{Q}_j = \sum_{j=1}^m \mathbf{A}_j^t \mathbf{A}_j \quad (2.59)$$

sendo as matrizes  $\mathbf{A}_j$  constituídas pelos vectores linha de  $\mathbf{P}$  com índices em  $\Upsilon_j$ , com  $j = 1, \dots, m$ . É fácil de verificar que as matrizes  $\mathbf{Q}_j$  são matrizes simétricas e idempotentes sendo, portanto, matrizes de projecção ortogonal e também mutuamente ortogonais.

Assim,  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$  será a base principal de uma Álgebra de Jordan Comutativa completa que tem  $\mathbf{P}$  como matriz ortogonal associada.

Sendo  $\mathbf{M} \in \mathcal{A}(\mathbf{P})$  uma matriz de projecção ortogonal, então atendendo a (2.53) se  $\text{car}(\mathbf{M}) = 1$ ,  $\mathbf{M}$  terá de pertencer à base principal  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$ .<sup>4</sup>

Se  $\mathcal{A}(\mathbf{P})$  for uma Álgebra de Jordan Comutativa regular com base principal  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$  vem que

$$\mathbf{Q}_1 = \frac{1}{n} \mathbf{J}_n \quad (2.60)$$

e ter-se-á

$$\mathbf{Q}_1 = \frac{1}{n} \mathbf{J}_n = \underbrace{\left( \frac{1}{\sqrt{n}} \mathbf{1}^n \right)}_{\mathbf{A}_1^t} \underbrace{\left( \frac{1}{\sqrt{n}} (\mathbf{1}^n)^t \right)}_{\mathbf{A}_1} \quad (2.61)$$

vindo

$$\mathbf{A}_1 = \left( \frac{1}{\sqrt{n}} (\mathbf{1}^n)^t \right), \quad (2.62)$$

em que  $\mathbf{1}^n$  é o vector coluna de ordem  $n$  cujos elementos também são todos iguais a 1.

Logo  $\mathbf{P}$  tem uma primeira linha com elementos iguais a  $\frac{1}{\sqrt{n}}$  e além disso, tem-se

$$\mathbf{A}_j \mathbf{A}_1^t = \mathbf{0}^{gj}, \text{ com } j = 2, \dots, m. \quad (2.63)$$

Como a soma de componentes de qualquer vector linha de  $\mathbf{A}_j$ , com  $j = 2, \dots, m$ , é nula, esses vectores serão vectores de contrastes. Portanto as matrizes  $\mathbf{A}_j$ , com  $j = 2, \dots, m$  serão matrizes de contrastes e  $\mathbf{P}$  será uma matriz ortogonal estandardizada, ver secção 2.1.5.

Seja  $\mathcal{F}(\mathbf{P})$  uma família de matrizes simétricas que têm como  $\mathbf{P}$  a sua matriz diagonalizadora ortogonal, então tem-se o seguinte resultado, ver [Ferreira, 2006]

**Proposição 2.2.4**  $\mathcal{F}(\mathbf{P})$  é uma Álgebra de Jordan Comutativa.

**Dem.**  $\mathcal{F}(\mathbf{P})$  é um espaço vectorial uma vez, que dadas as matrizes  $\mathbf{M}_1, \dots, \mathbf{M}_s \in \mathcal{F}(\mathbf{P})$ , uma sua combinação linear

$$\mathbf{M} = \sum_{j=1}^s b_j \mathbf{M}_j \in \mathcal{F}(\mathbf{P})$$

---

<sup>4</sup>Note-se que  $\frac{1}{n} \mathbf{J}_n$  será uma matriz de projecção ortogonal de característica 1, logo pertence à base principal de uma álgebra de Jordan comutativa, sempre que pertence à álgebra de Jordan comutativa.

pois

$$\mathbf{PMP}^t = \mathbf{P} \left( \sum_{j=1}^s b_j \mathbf{M}_j \right) \mathbf{P}^t = \sum_{j=1}^s b_j (\mathbf{PM}_j \mathbf{P}^t) = \sum_{j=1}^s b_j \mathbf{D}_j$$

continua a ser uma matriz diagonal e facilmente se verificam todas as outras condições para ser espaço vectorial.

Sabemos que as matrizes pertencentes a esta família comutam e utilizando a proposição 2.1.2, tem-se

$$(\mathbf{PMP}^t) (\mathbf{PMP}^t) = \mathbf{D}^2,$$

isto é,

$$\mathbf{PM}^2 \mathbf{P}^t = \mathbf{D}^2$$

é uma matriz diagonal, logo  $\mathbf{M}^2 \in \mathcal{F}(\mathbf{P})$ .

Sejam  $\alpha_1, \dots, \alpha_n$  os vectores linha de  $\mathbf{P}$  e  $r_1, \dots, r_n$  os valores próprios de  $\mathbf{M} \in \mathcal{F}(\mathbf{P})$ , então tem-se, pela decomposição espectral, ver [Schott, 1997], que

$$\mathbf{M} = \sum_{j=1}^n r_j \alpha_j \alpha_j^t.$$

Sendo  $\mathbf{M}$  uma matriz qualquer pertencente a  $\mathcal{F}(\mathbf{P})$  e que pode ser escrita como uma combinação linear não nula das matrizes  $\mathbf{Q}_j = \alpha_j \alpha_j^t$  linearmente independentes, então as matrizes  $\mathbf{Q}_j$ , com  $j = 1, \dots, n$  constituirão uma base para  $\mathcal{F}(\mathbf{P})$ .

Mostra-se ainda que sendo os vectores  $\alpha_1, \dots, \alpha_n$  vectores ortonormados, então as matrizes  $\mathbf{Q}_j$  são matrizes de projecção ortogonal (e portanto matrizes simétricas e idempotentes) e que são mutuamente ortogonais, ou seja,  $\mathbf{Q}_j \mathbf{Q}_i = \mathbf{O}_{n,n}$ , para  $j \neq i$  e  $j = 1, \dots, n$ . Assim sendo  $\mathcal{F}(\mathbf{P})$  tem dimensão  $n$ . ■

Falta ainda provar, relativamente à proposição anterior, que a base é única. Efectivamente, Seely em [Seely, 1971], chamou as Álgebras de Jordan Comutativas de espaços vectoriais comutativos quadráticos e apresentou o seguinte resultado

**Proposição 2.2.5** *Uma condição necessária e suficiente para um subespaço  $\mathcal{A}(\mathbf{P})$  ser Álgebra de Jordan Comutativa é a existência de uma base  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$  para esse subespaço, tal que cada matriz simétrica  $\mathbf{Q}_j$ , com  $j = 1, \dots, m$  é idempotente e  $\mathbf{Q}_i \mathbf{Q}_j = \mathbf{O}_{n,n}$ , para  $j \neq i$  e  $j = 1, \dots, m$ . Além disso, independentemente do índice das matrizes na base, essa base é única.*

**Dem.** A parte da condição necessária e suficiente é facilmente provada. A demonstração de que essa base é única apresenta-se de seguida.

Seja  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$  uma base de uma Álgebra de Jordan Comutativa  $\mathcal{A}(\mathbf{P})$ , constituída por matrizes de projecção ortogonal mutuamente ortogonais e suponha-se que se tinha outra base  $\{\mathbf{R}_1, \dots, \mathbf{R}_m\}$  para  $\mathcal{A}(\mathbf{P})$  formada por matrizes de projecção ortogonal e mutuamente ortogonais. Fixando  $w \in \{1, \dots, m\}$  sejam  $a_i$  números reais tais que

$$\mathbf{R}_w = \sum_{i=1}^m a_i \mathbf{Q}_i$$

e  $b_{h,j}$  números reais tais que

$$\mathbf{Q}_h = \sum_{j=1}^m b_{h,j} \mathbf{R}_j$$

para  $h = 1, \dots, m$ .

Então, uma vez que as matrizes  $\mathbf{Q}_i$ , para  $i = 1, \dots, m$  são mutuamente ortogonais e idempotentes, vem

$$\mathbf{R}_w \mathbf{Q}_h = \left( \sum_{i=1}^m a_i \mathbf{Q}_i \right) \mathbf{Q}_h = a_h \mathbf{Q}_h \mathbf{Q}_h = a_h \mathbf{Q}_h^2 = a_h \mathbf{Q}_h$$

o mesmo para matrizes  $\mathbf{R}_j$  para  $j = 1, \dots, m$

$$\mathbf{R}_w \mathbf{Q}_h = \mathbf{R}_w \left( \sum_{j=1}^m b_{h,j} \mathbf{R}_j \right) = \mathbf{R}_w b_{h,w} \mathbf{R}_w = b_{h,w} \mathbf{R}_w^2 = b_{h,w} \mathbf{R}_w$$

logo  $a_h \mathbf{Q}_h = b_{h,w} \mathbf{R}_w$  para  $h = 1, \dots, m$ . Como as matrizes  $\mathbf{Q}_i$  são linearmente independentes tem-se  $\mathbf{R}_w \neq 0$  e como  $\mathbf{Q}_h$  e  $\mathbf{R}_w$  são matrizes pertencentes a bases tem-se que  $(a_h = 0 \text{ ou } a_h = 1)$  e  $(b_{h,w} = 0 \text{ ou } b_{h,w} = 1)$ , e portanto  $\mathbf{R}_w = \mathbf{Q}_h$  para algum  $h$ . Dado que esta igualdade se verifica para  $w = 1, \dots, m$ , então está provada a unicidade da  $\mathbf{Q}_1, \dots, \mathbf{Q}_m$ . ■

## 2.3 Distribuições Associadas à Distribuição Normal

Existem várias distribuições relacionadas com a distribuição normal, mas dessas destacaremos a distribuição qui-quadrado e a distribuição  $F$  de Snedecor, em particular as distribuições qui-quadrado não central e  $F$  de Snedecor não central, muito utilizadas ao longo desta dissertação.

### 2.3.1 Distribuição Qui-quadrado Não Central

**Definição 2.3.1 (Distribuição Qui-Quadrado)** Sejam  $X_1, \dots, X_n$  variáveis aleatórias i.i.d.,  $X_i \sim N(0, 1)$ , logo  $\mathbf{X} = [X_1 \dots X_n]^t$  tem distribuição normal de vector médio  $\mathbf{0}^n$  e matriz de covariância  $\mathbf{I}_n$ . Assim

$$\sum_{i=1}^n X_i^2 = \mathbf{X}^t \mathbf{X} \sim \chi_n^2$$

isto é, a soma de quadrados de  $n$  variáveis aleatórias i.i.d.,  $X_i \sim N(0, 1)$ , tem distribuição qui-quadrado (central) com  $n$  graus de liberdade.

**Definição 2.3.2 (Distribuição Qui-Quadrado Não Central)** Sejam  $X_1, \dots, X_n$  variáveis aleatórias independentes cuja distribuição de  $X_i \sim N(\mu_i, 1)$ , de tal forma que  $\mathbf{X} = [X_1 \dots X_n]^t$  tem distribuição normal de vector médio  $\boldsymbol{\mu} = [\mu_1 \dots \mu_n]^t$  e matriz de covariância  $\mathbf{I}_n$ . Assim

$$\sum_{i=1}^n X_i^2 = \mathbf{X}^t \mathbf{X}$$

não tem distribuição qui-quadrado central, mas a soma de quadrados de desvios tem distribuição qui-quadrado central, ou seja,

$$\sum_{i=1}^n (X_i - \mu_i)^2 = (\mathbf{X} - \boldsymbol{\mu})^t (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_n^2,$$

uma vez que  $X_i - \mu_i \sim N(0, 1)$ .

Assim

$$\sum_{i=1}^n X_i^2 = \mathbf{X}^t \mathbf{X} \sim \chi_n^2(\delta),$$

onde  $\chi_n^2(\delta)$  é designada por distribuição qui-quadrado não central de parâmetro de não centralidade  $\delta$ , em que  $\delta$ <sup>5</sup> é definido como

$$\delta = \sum_{i=1}^n \mu_i^2 = \boldsymbol{\mu}^t \boldsymbol{\mu}.$$

**Teorema 2.3.1** Se  $Y_1, Y_2, \dots, Y_k$  são independentes cuja distribuição de  $Y_i \sim \chi_{n_i}^2(\delta_i)$  então

$$\sum_{i=1}^k Y_i \sim \chi_n^2(\delta)$$

---

<sup>5</sup>Note-se que em algumas referências o parâmetro de não centralidade é definido de outra forma, tal como metade da soma ou a sua raiz quadrada.



com

$$n = \sum_{i=1}^k n_i \quad e \quad \delta = \sum_{i=1}^k \delta_i.$$

**Corolário 2.3.1.1** Se  $Y_1, Y_2, \dots, Y_k$  são independentes cuja distribuição de  $Y_i \sim \chi_{n_i}^2$  então

$$\sum_{i=1}^k Y_i \sim \chi_n^2$$

com

$$n = \sum_{i=1}^k n_i.$$

### 2.3.2 Distribuição $F$ de Snedecor Não Central

**Definição 2.3.3 (Distribuição  $F$  de Snedecor)** Se  $X \sim \chi_p^2$ ,  $Y \sim \chi_q^2$  com  $X$  e  $Y$  independentes, então o quociente

$$W = \frac{\frac{X}{p}}{\frac{Y}{q}} \sim F_{p,q}$$

tem distribuição  $F$  de Snedecor (central) com  $p$  e  $q$  graus de liberdade.

**Definição 2.3.4 (Distribuição  $F$  de Snedecor Não Central)** Se  $X \sim \chi_p^2(\delta)$ ,  $Y \sim \chi_q^2$  com  $X$  e  $Y$  independentes, então o quociente

$$W = \frac{\frac{X}{p}}{\frac{Y}{q}} \sim F_{p,q}(\delta)$$

tem distribuição  $F$  de Snedecor não central com parâmetro de não centralidade  $\delta$ , onde  $\delta$  é o mesmo parâmetro de não centralidade da distribuição de  $X$ .

Quando uma estatística com distribuição  $F$  de Snedecor é usada para testar uma hipótese  $H_0$ , a distribuição será central se a hipótese  $H_0$  for verdadeira e terá distribuição  $F$  de Snedecor não central se essa hipótese for falsa. Assim, a distribuição  $F$  de Snedecor não central pode ser usada para avaliar a potência de um teste  $F$ .

### 2.3.3 Distribuição de Formas Quadráticas

**Teorema 2.3.2** *Seja  $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{I}_n)$  então*

$$(\mathbf{X} - \boldsymbol{\mu})^t (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_n^2.$$

**Teorema 2.3.3** *Seja  $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{C})$  então*

$$(\mathbf{X} - \boldsymbol{\mu})^t \mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_n^2.$$

**Teorema 2.3.4** *Sejam  $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{C})$ , onde  $\mathbf{C}$  é uma matriz definida positiva e  $\mathbf{A}$  uma matriz simétrica de constantes. Se  $\mathbf{AC}$  é idempotente e  $\text{car}(\mathbf{AC}) = r$  então*

$$\mathbf{X}^t \mathbf{A} \mathbf{X} \sim \chi_r^2(\lambda)$$

com

$$\lambda = \boldsymbol{\mu}^t \mathbf{A} \boldsymbol{\mu}.$$

**Dem.** Ver em [Schott, 1997]. ■

**Corolário 2.3.4.1** *Se  $\mathbf{X} \sim N(\mathbf{0}^n, \mathbf{I}_n)$  e  $\mathbf{A}$  é idempotente e  $\text{car}(\mathbf{A}) = r$ , então*

$$\mathbf{X}^t \mathbf{A} \mathbf{X} \sim \chi_r^2.$$

**Dem.** Ver em [Schott, 1997]. ■

**Corolário 2.3.4.2** *Se  $\mathbf{X} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$  e  $\mathbf{A}$  é idempotente e  $\text{car}(\mathbf{A}) = r$ , então*

$$\mathbf{X}^t \mathbf{A} \mathbf{X} \sim \sigma^2 \chi_r^2(\delta)$$

com

$$\delta = \frac{1}{\sigma^2} \boldsymbol{\mu}^t \mathbf{A} \boldsymbol{\mu}.$$

**Dem.** Ver em [Schott, 1997]. ■

## 2.4 Testes de Hipóteses - Testes $F$

- Seja  $\mathbf{Y}$  um vector aleatório tal que  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ . Então, ver [Mexia, 1995],  $\|\mathbf{Y}\|^2$  será o produto de  $\sigma^2$  por um qui-quadrado com  $n$  graus de liberdade e parâmetro de não centralidade dado por

$$\delta = \frac{1}{\sigma^2} \|\boldsymbol{\mu}\|^2 \tag{2.64}$$

escrevendo-se

$$\|\mathbf{Y}\|^2 \sim \sigma^2 \chi_n^2(\delta). \quad (2.65)$$

Consideremos

$$S \sim \sigma^2 \chi_m^2 \quad (2.66)$$

independente de  $\mathbf{Y}$ , então o quociente

$$\mathcal{F} = \frac{\frac{\|\mathbf{Y}\|^2}{n}}{\frac{S}{m}} \quad (2.67)$$

terá, ver [Mexia, 1995], distribuição  $F$  de Snedecor não central com  $n$  e  $m$  graus de liberdade e parâmetro de não centralidade  $\delta$ , escrevendo-se

$$\mathcal{F} \sim F_{n,m}(\delta). \quad (2.68)$$

- Em particular, sendo  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$  e se  $\boldsymbol{\mu} = \mathbf{0}^n$  então em (2.64) viria  $\delta = 0$ , e conseqüentemente

$$\|\mathbf{Y}\|^2 \sim \sigma^2 \chi_n^2 \quad (2.69)$$

e portanto  $\mathcal{F}$  terá distribuição  $F$  de Snedecor central, escrevendo-se

$$\mathcal{F} \sim F_{n,m}. \quad (2.70)$$

- Mais geralmente, ver [Mexia, 1995], se  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{M})$  com  $\text{car}(\mathbf{M}) = l \leq n$  tem-se

$$(\mathbf{Y} - \mathbf{b})^t \mathbf{M}^+ (\mathbf{Y} - \mathbf{b}) \sim \sigma^2 \chi_l^2(\delta) \quad (2.71)$$

sendo  $\mathbf{M}^+$  a matriz inversa de Moore-Penrose da matriz  $\mathbf{M}$  (uma vez que a inversa usual de  $\mathbf{M}$  não existe, pois  $\text{car}(\mathbf{M}) = l \leq n$ ) e o parâmetro de não centralidade dado por

$$\delta = \frac{1}{\sigma^2} (\boldsymbol{\mu} - \mathbf{b})^t \mathbf{M}^+ (\boldsymbol{\mu} - \mathbf{b}) \quad (2.72)$$

e em que  $\mathbf{b}$  é um vector qualquer.

Se considerarmos

$$S \sim \sigma^2 \chi_m^2 \quad (2.73)$$

e independente de  $\mathbf{Y}$ , ter-se-á

$$\mathcal{F} = \frac{m}{l} \frac{(\mathbf{Y} - \mathbf{b})^t \mathbf{M}^+ (\mathbf{Y} - \mathbf{b})}{S} \quad (2.74)$$

com distribuição  $F$  de Snedecor não central com  $l$  e  $m$  graus de liberdade e parâmetro de não centralidade  $\delta$ , escrevendo-se

$$\mathcal{F} \sim F_{l,m}(\delta). \quad (2.75)$$

- Em particular, sendo  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{M})$  e se  $\boldsymbol{\mu} = \mathbf{b}$  então em (2.72) viria  $\delta = 0$ , e conseqüentemente

$$(\mathbf{Y} - \mathbf{b})^t \mathbf{M}^+ (\mathbf{Y} - \mathbf{b}) \sim \sigma^2 \chi_l^2 \quad (2.76)$$

e

$$\mathcal{F}' = \frac{m (\mathbf{Y} - \mathbf{b})^t \mathbf{M}^+ (\mathbf{Y} - \mathbf{b})}{l S} \sim F_{l,m}, \quad (2.77)$$

ou seja  $\mathcal{F}'$  tem distribuição  $F$  de Snedecor central com  $l$  e  $m$  graus de liberdade.

Assim,  $\mathcal{F}$  será a estatística do teste utilizada para testar a hipótese

$$H_0 : \boldsymbol{\mu} = \mathbf{b}. \quad (2.78)$$

Designemos por  $f_{1-q,l,m}$  o quantil de probabilidade  $1 - q$  da distribuição  $F$  de Snedecor central com  $l$  e  $m$  graus de liberdade, ou seja

$$\Pr [\mathcal{F}' \leq f_{1-q,l,m}] = 1 - q. \quad (2.79)$$

Como  $\delta = 0$ , e por isso a hipótese  $H_0$  verifica-se, tem-se neste caso, que a potência do teste será dada por

$$Pot = 1 - \Pr [\mathcal{F}' \leq f_{1-q,l,m}] = \Pr [\mathcal{F}' > f_{1-q,l,m}] = q \quad (2.80)$$

donde concluímos que o teste será não distorcido.

Uma vez que

$$\begin{aligned} \Pr [\mathcal{F}' \leq f_{1-q,l,m}] &= 1 - q \Leftrightarrow \\ \Leftrightarrow \Pr \left[ \frac{m (\mathbf{Y} - \mathbf{b})^t \mathbf{M}^+ (\mathbf{Y} - \mathbf{b})}{l S} \leq f_{1-q,l,m} \right] &= 1 - q \Leftrightarrow \\ \Leftrightarrow \Pr \left[ (\mathbf{Y} - \mathbf{b})^t \mathbf{M}^+ (\mathbf{Y} - \mathbf{b}) \leq f_{1-q,l,m} \times \frac{l}{m} \times S \right] &= 1 - q \end{aligned}$$

como consequência, as desigualdades

$$(\mathbf{Y} - \boldsymbol{\mu})^t \mathbf{M}^+ (\mathbf{Y} - \boldsymbol{\mu}) \leq f_{1-q,l,m} \times \frac{l}{m} \times S \quad (2.81)$$

definem elipsóides de confiança de nível  $1 - q$  para  $\boldsymbol{\mu}$ , isto é, a probabilidade de  $\boldsymbol{\mu}$  estar coberto por este elipsóide é  $1 - q$ , donde o teste  $F$  de nível  $q$  não rejeita  $H_0$  se e só se  $\mathbf{b}$  pertencer ao elipsóide de confiança de nível  $1 - q$ , referido em (2.81). Diz-se, neste caso, que o teste goza de dualidade, ver [Mexia, 1990].

Quando  $\mathbf{M}$  é uma matriz definida positiva, significa que todos os seus valores próprios são todos positivos e consequentemente o determinante da matriz  $\mathbf{M}$  é diferente de 0, e por isso a matriz  $\mathbf{M}$  é invertível, logo  $\mathbf{M}^+ = \mathbf{M}^{-1}$  e  $\mathbf{M}^{-1}$  também é definida positiva, vindo

$$\delta = \frac{1}{\sigma^2} (\boldsymbol{\mu} - \mathbf{b})^t \mathbf{M}^{-1} (\boldsymbol{\mu} - \mathbf{b}) \quad (2.82)$$

e podendo reescrever-se a hipótese  $H_0 : \boldsymbol{\mu} = \mathbf{b}$  como  $H_0 : \delta = 0$ . Neste caso, o teste é estritamente não distorcido, ver [Mexia, 1995].



# Capítulo 3

## Modelos para matrizes estocásticas simétricas

### 3.1 Considerações Prévias

Este capítulo constitui a parte essencial da nossa dissertação. Nele apresentaremos modelos para matrizes estocásticas simétricas baseados na decomposição espectral das respectivas matrizes médias.

Assim, sendo  $\mathbf{S} = [S_{ij}]$  uma matriz estocástica simétrica, a respectiva matriz média  $E[\mathbf{S}] = [E(S_{ij})]$  será simétrica. A característica da matriz  $E[\mathbf{S}]$  será igual ao grau de  $\mathbf{S}$ . Se  $\mathbf{S}$  tem grau  $r$ , temos

$$E[\mathbf{S}] = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t, \quad (3.1)$$

facto este utilizado na secção seguinte para construir o modelo base.

### 3.2 Modelo Base

Em ([Oliveira e Mexia, 1998], [Oliveira e Mexia, 1999] e [Oliveira, 2001]) desenvolveu-se inferência estatística para matrizes com grau 1. Foram utilizados modelos da forma

$$\mathbf{S} = \lambda \boldsymbol{\alpha} \boldsymbol{\alpha}^t + \bar{\mathbf{E}} \quad (3.2)$$

em que  $\bar{\mathbf{E}} = \frac{1}{2}(\mathbf{E} + \mathbf{E}^t)$  sendo  $\mathbf{E} = [e_{ij}]$ ,  $i = 1, \dots, k$ ;  $j = 1, \dots, k$  com  $e_{ij}$  *i.i.d.* com distribuição normal de valor médio nulo e variância  $\sigma^2$ , ou seja,

$$e_{ij} \sim N(0, \sigma^2), \quad (3.3)$$

tendo

$$\text{vec}(\mathbf{E}) \sim N(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}), \quad (3.4)$$

em que  $\text{vec}(\mathbf{E})$  representa a vectorização da matriz  $\mathbf{E}$ .

No que se segue, alargaremos a formulação introduzida em [Oliveira e Mexia, 1999] e reformularemos o que foi apresentado em [Areia, 2009].

Vejamos, em primeiro lugar, como construir modelos da forma

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} \quad (3.5)$$

para matrizes com grau  $r \geq 1$  com  $\bar{\mathbf{E}} = \frac{1}{2}(\mathbf{E} + \mathbf{E}^t)$  sendo

$$\text{vec}(\mathbf{E}) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right). \quad (3.6)$$

Observe-se que

$$\boldsymbol{\alpha}_i^t \boldsymbol{\alpha}_j = \delta_{ij}, \text{ com } i = 1, \dots, r; j = 1, \dots, r, \quad (3.7)$$

em que  $\delta_{ij} = 0$  se  $i \neq j$  e  $\delta_{ij} = 1$  se  $i = j$ .

Considerando

$$\boldsymbol{\beta}_i = \lambda_i \boldsymbol{\alpha}_i, \text{ com } i = 1, \dots, r, \quad (3.8)$$

o modelo apresentado em (3.5), pode ser reescrito

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} = \sum_{i=1}^r \boldsymbol{\beta}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} \quad (3.9)$$

com  $\bar{\mathbf{E}} = [\bar{e}_{ij}]$  e  $\mathbf{E} = [e_{ij}]$ .

Repare-se que

$$\bar{e}_{ij} = \frac{1}{2}(e_{ij} + e_{ji}), i \neq j \quad \text{e} \quad \bar{e}_{ii} = \frac{1}{2}(e_{ii} + e_{ii}) = e_{ii}, \quad (3.10)$$

logo

$$\bar{e}_{ij} \sim N\left(0, \frac{\sigma^2}{2}\right), i \neq j \quad \text{e} \quad \bar{e}_{ii} \sim N(0, \sigma^2). \quad (3.11)$$

Se  $i \neq j$ , e sabendo que  $e_{ij} \sim N(0, \sigma^2)$ , tem-se

$$E(\bar{e}_{ij}) = E\left(\frac{1}{2}(e_{ij} + e_{ji})\right) = \frac{1}{2}(E(e_{ij}) + E(e_{ji})) = 0$$

e

$$\begin{aligned} V(\bar{e}_{ij}) &= V\left(\frac{1}{2}(e_{ij} + e_{ji})\right) = \frac{1}{4}(V(e_{ij}) + V(e_{ji})) = \\ &= \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2} \end{aligned}$$



logo

$$\bar{e}_{ij} \sim N\left(0, \frac{\sigma^2}{2}\right), \text{ para } i \neq j. \quad (3.12)$$

E se  $i = j$ , temos

$$E(\bar{e}_{ii}) = E\left(\frac{1}{2}(e_{ii} + e_{ii})\right) = E(e_{ii}) = 0$$

e

$$V(\bar{e}_{ii}) = V\left(\frac{1}{2}(e_{ii} + e_{ii})\right) = V(e_{ii}) = \sigma^2$$

logo

$$\bar{e}_{ii} \sim N(0, \sigma^2). \quad (3.13)$$

Por outro lado

$$Cov(\bar{e}_{ij}; \bar{e}_{ji}) \underset{\bar{E} \text{ é simétrica}}{=} Cov(\bar{e}_{ij}; \bar{e}_{ij}) = V(\bar{e}_{ij}) = \frac{\sigma^2}{2},$$

com  $i = 1, \dots, k$ ;  $j = 1, \dots, k$ , e  $i \neq j$  e

$$Cov(\bar{e}_{ii}; \bar{e}_{ii}) = V(\bar{e}_{ii}) = \sigma^2, \text{ com } i = 1, \dots, k; j = 1, \dots, k,$$

sendo nulas as covariâncias entre os restantes pares de elementos de  $\bar{\mathbf{E}}$ .

### 3.2.1 Validação do Modelo

O processo de validação de um modelo é uma fase bastante importante na construção de um modelo, pois poder-se-á utilizar um modelo não adequado. Assim, a validação do modelo permite-nos confirmar a adequação deste modelo aos dados.

#### 3.2.1.1 Operadores vectoriais

Vejamos alguns operadores vectoriais que facilitam a apresentação de resultados importantes inseridos neste capítulo e no seguinte.

Dada uma matriz simétrica  $\mathbf{A} = [a_{ij}]$  de ordem  $k$ , ou seja  $\mathbf{A} \in \Theta_k$  em que  $\Theta_k$  representa o conjunto de matrizes simétricas de ordem  $k$ , temos

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1(r-1)} & a_{1r} & a_{1(r+1)} & \cdots & a_{1(k-1)} & a_{1k} \\ a_{12} & a_{22} & \cdots & a_{2(r-1)} & a_{2r} & a_{2(r+1)} & \cdots & a_{2(k-1)} & a_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{1(r-1)} & a_{2(r-1)} & \cdots & a_{(r-1)(r-1)} & a_{(r-1)r} & a_{(r-1)(r+1)} & \cdots & a_{(r-1)(k-1)} & a_{(r-1)k} \\ a_{1r} & a_{2r} & \cdots & a_{(r-1)r} & a_{rr} & a_{r(r+1)} & \cdots & a_{r(k-1)} & a_{rk} \\ a_{1(r+1)} & a_{2(r+1)} & \cdots & a_{(r-1)(r+1)} & a_{r(r+1)} & a_{(r+1)(r+1)} & \cdots & a_{(r+1)(k-1)} & a_{(r+1)k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{1(k-1)} & a_{2(k-1)} & \cdots & a_{(r-1)(k-1)} & a_{r(k-1)} & a_{(r+1)(k-1)} & \cdots & a_{(k-1)(k-1)} & a_{(k-1)k} \\ a_{1k} & a_{2k} & \cdots & a_{(r-1)k} & a_{rk} & a_{(r+1)k} & \cdots & a_{(k-1)k} & a_{kk} \end{bmatrix}$$

e representaremos por  $s\_vec(\mathbf{A})$  o vector columna

$$s\_vec(\mathbf{A}) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{rr} \\ \vdots \\ a_{kk} \\ \sqrt{2}a_{12} \\ \vdots \\ \sqrt{2}a_{(k-1)k} \end{bmatrix} \quad (3.14)$$

do tipo  $g \times 1$  com  $g = \frac{k(k+1)}{2}$ , tendo-se

$$\|\mathbf{A}\|^2 = \|s\_vec(\mathbf{A})\|^2. \quad (3.15)$$

Consideremos  $a\_vec(\mathbf{A}; r)$  e  $b\_vec(\mathbf{A}; r)$  dados por

$$a\_vec(\mathbf{A}; r) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{rr} \\ \sqrt{2}a_{12} \\ \vdots \\ \sqrt{2}a_{rk} \end{bmatrix}; \quad b\_vec(\mathbf{A}; r) = \begin{bmatrix} a_{(r+1)(r+1)} \\ \vdots \\ a_{kk} \\ \sqrt{2}a_{(r+1)(r+2)} \\ \vdots \\ \sqrt{2}a_{(k-1)k} \end{bmatrix} \quad (3.16)$$

do tipo  $a_r \times 1$  e  $b_r \times 1$ , respectivamente, com  $a_r = g - b_r$  e  $b_r = \frac{(k-r)(k-r+1)}{2}$ .

Note-se ainda que estes operadores gozam das seguintes propriedades, fáceis de estabelecer.

- $s\_vec \left( \sum_{l=1}^m c_l \mathbf{A}_l \right) = \sum_{l=1}^m (c_l \times s\_vec(\mathbf{A}_l))$ ;
- $a\_vec \left( \sum_{l=1}^m c_l \mathbf{A}_l ; r \right) = \sum_{l=1}^m (c_l \times a\_vec(\mathbf{A}_l ; r))$ ;
- $b\_vec \left( \sum_{l=1}^m c_l \mathbf{A}_l ; r \right) = \sum_{l=1}^m (c_l \times b\_vec(\mathbf{A}_l ; r))$ ;
- $\|\mathbf{A}\|^2 = \|a\_vec(\mathbf{A}; r)\|^2 + \|b\_vec(\mathbf{A}; r)\|^2$ .

Note-se que esta última propriedade pode ser apresentada por

$$\|s\_vec(\mathbf{A})\|^2 = \|a\_vec(\mathbf{A}; r)\|^2 + \|b\_vec(\mathbf{A}; r)\|^2 \quad (3.17)$$

atendendo a (3.15).

### 3.2.1.2 Resultados importantes

Interessa-nos agora estabelecer os seguintes lema e proposição

**Lema 3.2.1** *Sendo  $\mathbf{P}$  uma matriz ortogonal e  $vec(\mathbf{E}) \sim N(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2})$ , tem-se  $vec(\mathbf{PE}) \sim N(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2})$ .*

**Dem.** Como

$$vec(\mathbf{E}) \sim N(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2})$$

com  $\mathbf{E} = [\mathbf{e}_1 \dots \mathbf{e}_k]$  tem-se  $\mathbf{e}_1, \dots, \mathbf{e}_k$  *i.i.d.* com distribuição  $N(\mathbf{0}^k, \sigma^2 \mathbf{I}_k)$ , vindo  $\mathbf{PE} = [\mathbf{Pe}_1 \dots \mathbf{Pe}_k]$  com  $\mathbf{Pe}_1, \dots, \mathbf{Pe}_k$  *i.i.d.* e com distribuição  $N(\mathbf{0}^k, \sigma^2 \mathbf{I}_k)$  logo

$$vec(\mathbf{PE}) \sim N(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}),$$

o que estabelece a tese. ■

**Proposição 3.2.1** *Seja  $\mathbf{P}$  uma qualquer matriz ortogonal de ordem  $k$ . Então*

$$s\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t) \sim N(\mathbf{0}^g, \sigma^2 \mathbf{I}_g),$$

$$a\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r) \sim N(\mathbf{0}^{a_r}, \sigma^2 \mathbf{I}_{a_r}),$$

e

$$b\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r) \sim N(\mathbf{0}^{b_r}, \sigma^2 \mathbf{I}_{b_r}),$$

com  $g = \frac{k(k+1)}{2}$ ,  $a_r = g - \frac{(k-r)(k-r+1)}{2} = g - b_r$  e  $b_r = \frac{(k-r)(k-r+1)}{2}$ .

Note-se ainda que

$$\|s\_ \text{vec}(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t)\|^2 \sim \sigma^2 \chi_g^2,$$

$$\|a\_ \text{vec}(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r)\|^2 \sim \sigma^2 \chi_{a_r}^2,$$

e

$$\|b\_ \text{vec}(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r)\|^2 \sim \sigma^2 \chi_{b_r}^2.$$

**Dem.** De acordo com o lema 3.2.1, tem-se

$$\text{vec}(\mathbf{P}\mathbf{E}) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right)$$

e dado  $\text{vec}(\mathbf{E}^t \mathbf{P}^t) = \text{vec}((\mathbf{P}\mathbf{E})^t)$  ter as mesmas componentes de  $\text{vec}(\mathbf{P}\mathbf{E})$ , também temos

$$\text{vec}(\mathbf{E}^t \mathbf{P}^t) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right).$$

Sabemos também que

$$\text{vec}(\mathbf{E}) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right),$$

logo  $\text{vec}(\mathbf{E}^t)$  terá as mesmas componentes que  $\text{vec}(\mathbf{E})$ , e portanto

$$\text{vec}(\mathbf{E}^t) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right).$$

Assim, podemos também dizer que

$$\text{vec}\left((\mathbf{E}^t)^t \mathbf{P}^t\right) = \text{vec}(\mathbf{E}\mathbf{P}^t) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right),$$

e finalmente, podemos agora voltar a utilizar o lema 3.2.1, que atendendo ao facto de  $\mathbf{P}$  ser uma matriz ortogonal temos

$$\text{vec}(\mathbf{P}\mathbf{E}\mathbf{P}^t) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right).$$

Podemos também dizer que sendo

$$\text{vec}(\mathbf{E}^t \mathbf{P}^t) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right),$$

e voltando a utilizar o lema 3.2.1, temos

$$\text{vec}(\mathbf{P}\mathbf{E}^t \mathbf{P}^t) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right).$$

Consideremos então  $\mathbf{PEP}^t = [e_{ij}^\circ]$  e sendo  $(\mathbf{PE}^t \mathbf{P}^t)^t = \mathbf{PEP}^t$  então

$$\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t = \mathbf{P} \left( \frac{1}{2} (\mathbf{E} + \mathbf{E}^t) \right) \mathbf{P}^t = \frac{1}{2} \mathbf{PEP}^t + \frac{1}{2} \mathbf{PE}^t \mathbf{P}^t = \frac{1}{2} (\mathbf{PEP}^t + \mathbf{PE}^t \mathbf{P}^t),$$

logo as componentes de  $s\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t)$  serão

$$e_{11}^\circ, e_{22}^\circ, \dots, e_{rr}^\circ, \dots, e_{kk}^\circ, \sqrt{2} \left( \frac{1}{2} (e_{12}^\circ + e_{21}^\circ) \right), \sqrt{2} \left( \frac{1}{2} (e_{13}^\circ + e_{31}^\circ) \right), \dots, \sqrt{2} \left( \frac{1}{2} (e_{(k-1)k}^\circ + e_{k(k-1)}^\circ) \right),$$

sendo as componentes de  $a\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r)$

$$e_{11}^\circ, e_{22}^\circ, \dots, e_{rr}^\circ, \sqrt{2} \left( \frac{1}{2} (e_{12}^\circ + e_{21}^\circ) \right), \sqrt{2} \left( \frac{1}{2} (e_{13}^\circ + e_{31}^\circ) \right), \dots, \sqrt{2} \left( \frac{1}{2} (e_{rk}^\circ + e_{kr}^\circ) \right),$$

e as componentes de  $b\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r)$

$$e_{(r+1)(r+1)}^\circ, \dots, e_{kk}^\circ, \sqrt{2} \left( \frac{1}{2} (e_{(r+1)(r+2)}^\circ + e_{(r+2)(r+1)}^\circ) \right), \dots, \sqrt{2} \left( \frac{1}{2} (e_{(k-1)k}^\circ + e_{k(k-1)}^\circ) \right).$$

Note-se ainda que  $s\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t)$  tem  $g$  componentes com

$$g = k + \frac{k^2 - k}{2} = \frac{2k + k^2 - k}{2} = \frac{k^2 + k}{2} = \frac{k(k+1)}{2},$$

$b\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r)$  tem  $b_r$  componentes com

$$\begin{aligned} b_r &= (k-r) + \frac{(k-r)^2 - (k-r)}{2} = \frac{2(k-r) + (k-r)^2 - (k-r)}{2} = \\ &= \frac{(k-r)(k-r+1)}{2}, \end{aligned}$$

e  $a\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r)$  tem  $a_r$  componentes com

$$a_r = g - b_r.$$

Atendendo ao facto de

$$vec(\mathbf{PEP}^t) \sim N(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2})$$

e

$$vec(\mathbf{PE}^t \mathbf{P}^t) \sim N(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}),$$

vem que

$$s\_vec(\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t) \sim N(\mathbf{0}^g, \sigma^2 \mathbf{I}_g).$$

Podemos também concluir que

$$a\_ \text{vec} (\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t) \sim N (\mathbf{0}^{a_r}, \sigma^2 \mathbf{I}_{a_r}),$$

e

$$b\_ \text{vec} (\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t) \sim N (\mathbf{0}^{b_r}, \sigma^2 \mathbf{I}_{b_r}).$$

Consequentemente temos

$$\|s\_ \text{vec} (\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t)\|^2 \sim \sigma^2 \chi_g^2,$$

$$\|a\_ \text{vec} (\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r)\|^2 \sim \sigma^2 \chi_{a_r}^2,$$

e

$$\|b\_ \text{vec} (\mathbf{P}\bar{\mathbf{E}}\mathbf{P}^t; r)\|^2 \sim \sigma^2 \chi_{b_r}^2,$$

o que completa a demonstração. ■

### 3.2.2 Grau do modelo

Sejam  $\theta_1, \dots, \theta_k$  os valores próprios associados aos vectores próprios  $\gamma_1, \dots, \gamma_k$  da matriz  $\mathbf{S}$  cuja decomposição espectral é

$$\mathbf{S} = \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t. \quad (3.18)$$

Pode-se então escolher matrizes ortogonais  $\mathbf{P}^*$  e  $\mathbf{P}^\circ$  tais que

$$\mathbf{P}^* = \begin{bmatrix} \gamma_1^t \\ \vdots \\ \gamma_k^t \end{bmatrix} \quad \text{e} \quad \mathbf{P}^\circ = \begin{bmatrix} \alpha_1^t \\ \vdots \\ \alpha_k^t \end{bmatrix} \quad (3.19)$$

e admitindo que  $\gamma_i \approx \alpha_i$  e  $\theta_i \approx \lambda_i$  para  $i = 1, \dots, r$ , sendo  $\mathbf{v}_i^k(\mathbf{A})$  o  $i$ -ésimo vector linha de uma matriz  $\mathbf{A}$ , temos  $\mathbf{v}_i^k(\mathbf{P}^*) \approx \mathbf{v}_i^k(\mathbf{P}^\circ)$ , para  $i = 1, \dots, r$ , ver [Vaquinhas e Mexia, 1992].

Por outro lado, temos

$$\begin{aligned} \mathbf{P}^* \left( \sum_{i=1}^r \lambda_i \alpha_i \alpha_i^t \right) (\mathbf{P}^*)^t &\approx \mathbf{P}^* \left( \sum_{i=1}^r \theta_i \gamma_i \gamma_i^t \right) (\mathbf{P}^*)^t = \\ &= \sum_{i=1}^r (\theta_i \mathbf{P}^* \gamma_i \gamma_i^t (\mathbf{P}^*)^t) = \sum_{i=1}^r (\theta_i (\mathbf{P}^* \gamma_i) (\mathbf{P}^* \gamma_i)^t) = \\ &= \sum_{i=1}^r \left( \theta_i \delta_i^k (\delta_i^k)^t \right) = \mathbf{D} (\theta_1, \dots, \theta_r, 0, \dots, 0) \end{aligned}$$

sendo  $\delta_i^k$  o vector com a  $i$ -ésima componente igual a 1 e as restantes nulas e  $\mathbf{D}(\theta_1, \dots, \theta_r, 0, \dots, 0)$  representa a matriz diagonal cujos  $r$  primeiros elementos da diagonal são  $\theta_1, \dots, \theta_r$ , respectivamente e os restantes  $k - r$  elementos da diagonal são nulos. Assim, concluimos, atendendo ao modelo (3.9), que

$$\begin{aligned} \mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t &= \mathbf{P}^* \left( \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} \right) (\mathbf{P}^*)^t = \\ &= \mathbf{P}^* \left( \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t \right) (\mathbf{P}^*)^t + \mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t \approx \\ &\approx \mathbf{D}(\theta_1, \dots, \theta_r, 0, \dots, 0) + \mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t \end{aligned}$$

isto é

$$\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t \approx \mathbf{D}(\theta_1, \dots, \theta_r, 0, \dots, 0) + \mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t \quad (3.20)$$

Assim, atendendo a (3.20) temos

$$\begin{aligned} b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r) &\approx b\_vec(\mathbf{D}(\theta_1, \dots, \theta_r, 0, \dots, 0) + \mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r) = \\ &= b\_vec(\mathbf{D}(\theta_1, \dots, \theta_r, 0, \dots, 0) ; r) + b\_vec(\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r) = \\ &= \mathbf{0}^{b_r} + b\_vec(\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r) = b\_vec(\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r) \end{aligned}$$

e uma vez que através da proposição 3.2.1, temos

$$b\_vec(\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r) \sim N(\mathbf{0}^{b_r}, \sigma^2 \mathbf{I}_{b_r}). \quad (3.21)$$

então temos também que

$$b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r) \rightsquigarrow N(\mathbf{0}^{b_r}, \sigma^2 \mathbf{I}_{b_r}), \quad (3.22)$$

onde  $\rightsquigarrow$  indica distribuição aproximada.

Consequentemente podemos dizer que

$$\|b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2 \rightsquigarrow \sigma^2 \chi_{b_r}^2. \quad (3.23)$$

Em relação ao operador  $a\_vec(\cdot)$ , temos ainda que

$$a\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r) \approx a\_vec(\mathbf{D}(\theta_1, \dots, \theta_r, 0, \dots, 0) + \mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r) =$$

$$\begin{aligned}
&= a\_ \text{vec} (\mathbf{D} (\theta_1, \dots, \theta_r, 0, \dots, 0) ; r) + a\_ \text{vec} (\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r) = \\
&= \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} + a\_ \text{vec} (\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r) = \\
&= \underline{\boldsymbol{\theta}}_r + a\_ \text{vec} (\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r),
\end{aligned}$$

onde  $\theta_1, \dots, \theta_r$  são os  $r$  primeiros valores próprios da matriz  $S$ . E uma vez que

$$a\_ \text{vec} (\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t ; r) \sim N (\mathbf{0}^{a_r}, \sigma^2 \mathbf{I}_{a_r}), \quad (3.24)$$

como sabemos pela proposição 3.2.1, então concluimos que

$$a\_ \text{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r) \rightsquigarrow N (\underline{\boldsymbol{\theta}}_r, \sigma^2 \mathbf{I}_{a_r}). \quad (3.25)$$

Assim por (3.25), temos

$$\|a\_ \text{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2 \rightsquigarrow \sigma^2 \chi_{a_r}^2 (\delta_r) \quad (3.26)$$

com

$$\delta_r = \frac{1}{\sigma^2} \underline{\boldsymbol{\theta}}_r^t \underline{\boldsymbol{\theta}}_r = \frac{1}{\sigma^2} \|\underline{\boldsymbol{\theta}}_r\|^2 = \frac{1}{\sigma^2} \sum_{i=1}^r \theta_i^2 \approx \frac{1}{\sigma^2} \sum_{i=1}^r \lambda_i^2. \quad (3.27)$$

Atendendo que  $\|a\_ \text{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2$  é independente de  $\|b\_ \text{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2$ , e por (3.23) e (3.26) temos

$$\mathcal{F}_r = \frac{\frac{\|a\_ \text{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2}{a_r}}{\frac{\|b\_ \text{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2}{b_r}} = \frac{b_r}{a_r} \frac{\|a\_ \text{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2}{\|b\_ \text{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2} \quad (3.28)$$

que tem aproximadamente distribuição  $F$  não central com  $a_r$  e  $b_r$  graus de liberdade e parâmetro de não centralidade  $\delta_r$ , escrevendo-se

$$\mathcal{F}_r \rightsquigarrow F_{a_r, b_r} (\delta_r), \quad (3.29)$$

com

$$\delta_r = \frac{1}{\sigma^2} \sum_{i=1}^r \theta_i^2 \approx \frac{1}{\sigma^2} \sum_{i=1}^r \lambda_i^2. \quad (3.30)$$



Consideremos  $Y_1, \dots, Y_{b_r}$  as componentes de  $b\_ \text{vec}(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)$ . Quando o modelo (3.9) se ajusta somos levados a não rejeitar a hipótese

$$H_0 : Y_1, \dots, Y_{b_r} \text{ são } i.i.d. \text{ com distribuição aproximadamente } N(0, \sigma^2) \quad (3.31)$$

contra a hipótese alternativa

$$H_1 : Y_1, \dots, Y_{b_r} \text{ são independentes mas tais que } Y_i \rightsquigarrow N(\mu_i, \sigma^2), \quad (3.32)$$

com  $i = 1, \dots, b_r$ .

Para testar a hipótese  $H_0$  em (3.31) temos, ver [Mexia, 1995], a estatística de teste dada por

$$\overline{\mathcal{F}} = \frac{\left( \sum_{i=1}^{b_r} Y_i \right)^2}{b_r \sum_{i=1}^{b_r} Y_i^2 - \left( \sum_{i=1}^{b_r} Y_i \right)^2} \quad (3.33)$$

que tem, caso  $H_0$  se verifique, a distribuição do quociente de dois qui-quadrados centrais e independentes,  $\chi_1^2$  e  $\chi_{b_r-1}^2$ , ou seja,  $\overline{\mathcal{F}} \sim \overline{F}_{1, b_r-1}$ .

De facto, esta estatística de teste foi construída para testes cuja hipótese nula é  $H_0$  referida em (3.31) contra a hipótese alternativa em que as variáveis  $Y_1, \dots, Y_{b_r}$  são independentes mas não identicamente distribuídas, ou seja,  $H_1$  referida em (3.32).

Para considerar o que se passa quando a hipótese  $H_0$  referida em (3.31) não é verdadeira, tomemos

$$\mu_0 = \sum_{i=1}^{b_r} \mu_i \quad (3.34)$$

Neste caso, o numerador e o denominador de  $\overline{\mathcal{F}}$ , podem-se exprimir como o quociente de dois qui-quadrados independentes, mas não centrais, com 1 e  $b_r - 1$  graus de liberdade e parâmetros de não centralidade  $\delta_1$  e  $\delta_2$ , respectivamente, ver [Mexia, 1989a], isto é

$$\overline{\mathcal{F}} \sim \frac{\chi_1^2(\delta_1)}{\chi_{b_r-1}^2(\delta_2)} \quad (3.35)$$

e em que

$$\delta_1 = \frac{1}{b_r \sigma^2} \left( \sum_{i=1}^{b_r} \mu_i \right)^2 = \frac{\mu_0^2}{b_r \sigma^2} \quad (3.36)$$

e

$$\delta_2 = \frac{1}{\sigma^2} \left( \sum_{i=1}^{b_r} \mu_i^2 - \frac{1}{b_r} \left( \sum_{i=1}^{b_r} \mu_i \right)^2 \right) = \frac{1}{\sigma^2} \left( \sum_{i=1}^{b_r} \mu_i^2 - \frac{\mu_0^2}{b_r} \right). \quad (3.37)$$

Repare-se que se

$$Y_i \sim N(\mu_i, \sigma^2)$$

então

$$\sum_{i=1}^{b_r} Y_i \sim N\left(\sum_{i=1}^{b_r} \mu_i, b_r \sigma^2\right)$$

e

$$\frac{\sum_{i=1}^{b_r} Y_i}{\sqrt{b_r \sigma^2}} \sim N\left(\sum_{i=1}^{b_r} \mu_i, 1\right)$$

logo

$$\left(\frac{\sum_{i=1}^{b_r} Y_i}{\sqrt{b_r \sigma^2}}\right)^2 = \frac{1}{b_r \sigma^2} \left(\sum_{i=1}^{b_r} Y_i\right)^2 \sim \chi_1^2(\delta_1) \quad (3.38)$$

com

$$\delta_1 = \frac{1}{b_r \sigma^2} \left(\sum_{i=1}^{b_r} \mu_i\right)^2.$$

E sendo

$$Y_i \sim N(\mu_i, \sigma^2)$$

então

$$\frac{Y_i}{\sigma} \sim N(\mu_i, 1),$$

logo

$$\sum_{i=1}^{b_r} \left(\frac{Y_i}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{b_r} Y_i^2 \sim \chi_{b_r}^2(\delta)$$

com

$$\delta = \frac{1}{\sigma^2} \sum_{i=1}^{b_r} \mu_i^2.$$

Ora

$$\frac{1}{\sigma^2} \sum_{i=1}^{b_r} Y_i^2 - \frac{1}{b_r \sigma^2} \left(\sum_{i=1}^{b_r} Y_i\right)^2 \sim \chi_{b_r-1}^2(\delta_2) \quad (3.39)$$

com

$$\begin{aligned} \delta_2 &= \delta - \delta_1 = \frac{1}{\sigma^2} \sum_{i=1}^{b_r} \mu_i^2 - \frac{1}{b_r \sigma^2} \left(\sum_{i=1}^{b_r} \mu_i\right)^2 = \\ &= \frac{1}{\sigma^2} \left(\sum_{i=1}^{b_r} \mu_i^2 - \frac{1}{b_r} \left(\sum_{i=1}^{b_r} \mu_i\right)^2\right) \end{aligned}$$

Assim, por (3.38) e (3.39) vem

$$\begin{aligned}\overline{\mathcal{F}} &= \frac{\frac{1}{b_r \sigma^2} \left( \sum_{i=1}^{b_r} Y_i \right)^2}{\frac{1}{\sigma^2} \sum_{i=1}^{b_r} Y_i^2 - \frac{1}{b_r \sigma^2} \left( \sum_{i=1}^{b_r} Y_i \right)^2} = \frac{\frac{1}{b_r \sigma^2} \left( \sum_{i=1}^{b_r} Y_i \right)^2}{\frac{1}{b_r \sigma^2} \left( b_r \sum_{i=1}^{b_r} Y_i^2 - \left( \sum_{i=1}^{b_r} Y_i \right)^2 \right)} = \\ &= \frac{\left( \sum_{i=1}^{b_r} Y_i \right)^2}{b_r \sum_{i=1}^{b_r} Y_i^2 - \left( \sum_{i=1}^{b_r} Y_i \right)^2}\end{aligned}$$

Observe-se que  $H_0$  referida em (3.31), verifica-se se e só se  $\delta_1 = \delta_2 = 0$ .

Existem alternativas em que  $\delta_1$  predomina sobre  $\delta_2$  e outras em que  $\delta_2$  predomina sobre  $\delta_1$ .

Quando  $\delta_1$  predomina sobre  $\delta_2$  :

- a estatística  $\overline{\mathcal{F}}$  tende a tomar valores superiores aos que tomaria quando  $H_0$  se verifica.

Quando  $\delta_2$  predomina sobre  $\delta_1$  :

- a estatística  $\overline{\mathcal{F}}$  tende a tomar valores inferiores aos que tomaria quando  $H_0$  se verifica.

Assim, segundo [Mexia, 1989a], podemos considerar estes testes como sendo testes bilaterais.

Considerando  $\overline{f}_{p,1,b_r-1}$  o  $p$ -ésimo quantil da distribuição  $\overline{F}_{1,b_r-1}$ , temos a região de não rejeição do teste para um nível  $q$  dada por

$$\left[ \overline{f}_{\frac{q}{2},1,b_r-1}; \overline{f}_{1-\frac{q}{2},1,b_r-1} \right]. \quad (3.40)$$

Verifica-se que para os níveis de significância usuais, a potência destes testes cresce rapidamente com  $\delta_1$  e  $\delta_2$ , ver [Mexia, 1989a].

Observe-se ainda, por (3.36) e (3.37), que

$$\begin{aligned}
 \delta_1 + \delta_2 &= \frac{1}{b_r \sigma^2} \left( \sum_{i=1}^{b_r} \mu_i \right)^2 + \frac{1}{\sigma^2} \left( \sum_{i=1}^{b_r} \mu_i^2 - \frac{1}{b_r} \left( \sum_{i=1}^{b_r} \mu_i \right)^2 \right) = \\
 &= \frac{1}{b_r \sigma^2} \left( \sum_{i=1}^{b_r} \mu_i \right)^2 + \frac{1}{\sigma^2} \sum_{i=1}^{b_r} \mu_i^2 - \frac{1}{b_r \sigma^2} \left( \sum_{i=1}^{b_r} \mu_i \right)^2 = \\
 &= \frac{1}{\sigma^2} \sum_{i=1}^{b_r} \mu_i^2.
 \end{aligned}$$

Note-se ainda, que no caso em que se está a considerar a rejeição de  $H_0$ , esta pode resultar de

- não adequação do modelo (3.9);
- $\theta_i$  e  $\gamma_i$  não serem bons estimadores de  $\lambda_i$  e  $\alpha_i$ , respectivamente, com  $i = 1, \dots, r$ .

Logo, quando  $H_0$  não é rejeitada, somos levados a admitir que

- o modelo (3.9) ajusta-se;
- $\theta_i$  e  $\gamma_i$  são bons estimadores de  $\lambda_i$  e  $\alpha_i$ , respectivamente, com  $i = 1, \dots, r$ .

Observe-se que o modelo (3.9) poder-se-ia ajustar com  $\underline{\lambda}^r = \mathbf{0}^r$ , sendo

$$\underline{\lambda}^r = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_r \end{bmatrix}, \quad (3.41)$$

reduzindo-se o modelo (3.9) ao modelo  $\mathbf{S} = \overline{\mathbf{E}}$ , sendo este modelo pouco interessante do ponto de vista das aplicações.

Interessa-nos pois efectuar o seguinte teste de hipóteses

$$\begin{cases} H_{0,r} : \|\underline{\lambda}^r\| = 0 \\ \quad \quad \quad vs \\ H_{1,r} : \|\underline{\lambda}^r\| > 0 \end{cases} \quad (3.42)$$

Estabeleça-se a proposição

**Proposição 3.2.2** Quando no teste (3.42),  $H_{0,r}$  não se verifica, a estatística de teste

$$\mathcal{F}_r = \frac{b_r \left\| a_- \text{vec} \left( \mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r \right) \right\|^2}{a_r \left\| b_- \text{vec} \left( \mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r \right) \right\|^2}$$

tem aproximadamente distribuição  $F$  não central com  $a_r$  e  $b_r$  graus de liberdade e parâmetro de não centralidade  $\delta_r$ , com

$$\delta_r = \frac{1}{\sigma^2} \sum_{i=1}^r \theta_i^2 \approx \frac{1}{\sigma^2} \sum_{i=1}^r \lambda_i^2$$

escrevendo-se

$$\mathcal{F}_r \rightsquigarrow F_{a_r, b_r}(\delta_r).$$

**Dem.** Raciocinando-se como atrás foi feito, demonstra-se esta proposição. ■

Observe-se ainda que o modelo (3.9) poder-se-ia simplificar caso o valor próprio  $\lambda_r$  fosse igual a 0, reduzindo-se o modelo (3.9) ao modelo

$$\mathbf{S} = \sum_{i=1}^{r-1} \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} = \sum_{i=1}^{r-1} \beta_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}}. \quad (3.43)$$

Este último modelo seria um modelo com grau  $r - 1$ , sendo  $r - 1$  a característica da componente determinística do modelo, ou seja,  $\sum_{i=1}^{r-1} \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t$  teria característica  $r - 1$ .

Assim, com este propósito, interessa-nos efectuar também o teste de hipóteses

$$\begin{cases} H_{0,r}^o : \lambda_r = 0 \\ \quad \quad \quad vs \\ H_{1,r}^o : \lambda_r \neq 0 \end{cases} \quad (3.44)$$

para verificar o grau do modelo.

Consideremos a matriz  $\mathbf{U} = [u_{ij}]$  tal que

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1j} & u_{1(j+1)} & \cdots & u_{1(k-1)} & u_{1k} \\ u_{21} & u_{22} & \cdots & u_{2j} & u_{2(j+1)} & \cdots & u_{2(k-1)} & u_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{j1} & u_{j2} & \cdots & u_{jj} & u_{j(j+1)} & \cdots & u_{j(k-1)} & u_{jk} \\ u_{(j+1)1} & u_{(j+1)2} & \cdots & u_{(j+1)j} & u_{(j+1)(j+1)} & \cdots & u_{(j+1)(k-1)} & u_{(j+1)k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ u_{(k-1)1} & u_{(k-1)2} & \cdots & u_{(k-1)j} & u_{(k-1)(j+1)} & \cdots & u_{(k-1)(k-1)} & u_{(k-1)k} \\ u_{k1} & u_{k2} & \cdots & u_{kj} & u_{k(j+1)} & \cdots & u_{k(k-1)} & u_{kk} \end{bmatrix}$$

e seja

$$\underline{\mathbf{v}}_j(\mathbf{U}) = \begin{bmatrix} u_{jj} \\ u_{j(j+1)} \\ \vdots \\ u_{j(k-1)} \\ u_{jk} \end{bmatrix} \quad (3.45)$$

o  $j$ -ésimo vector linha da matriz  $\mathbf{U}$  a começar na posição  $(j, j)$ . O vector  $\underline{\mathbf{v}}_j(\mathbf{U})$  é um vector com  $c_j$  componentes, em que  $c_j = k - j + 1$ .

Admitindo que  $\theta_i \approx \lambda_i$ , para  $i = 1, \dots, r$ , temos

$$\begin{aligned} \underline{\mathbf{v}}_r(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t) &= \underline{\mathbf{v}}_r(\mathbf{D}(\theta_1, \dots, \theta_r, 0, \dots, 0) + \mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t) \approx \\ &\approx \underline{\mathbf{v}}_r(\mathbf{D}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) + \mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t) = \\ &= \underline{\mathbf{v}}_r(\mathbf{D}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)) + \underline{\mathbf{v}}_r(\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t) = \\ &= \begin{bmatrix} \lambda_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \underline{\mathbf{v}}_r(\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t). \end{aligned}$$

Assim, uma vez que

$$\underline{\mathbf{v}}_r(\mathbf{P}^* \bar{\mathbf{E}} (\mathbf{P}^*)^t) \sim N(\mathbf{0}^{k-r+1}, \sigma^2 \mathbf{I}_{k-r+1}), \quad (3.46)$$

podemos dizer que

$$\underline{\mathbf{v}}_r(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t) \sim N(\lambda_r \boldsymbol{\delta}_1^{k-r+1}, \sigma^2 \mathbf{I}_{k-r+1}), \quad (3.47)$$

sendo  $\boldsymbol{\delta}_1^{k-r+1}$  o vector coluna do tipo  $(k - r + 1) \times 1$  com a primeira componente igual a 1 e as restantes nulas. Consequentemente temos

$$\|\underline{\mathbf{v}}_r(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t)\|^2 \sim \sigma^2 \chi_{k-r+1}^2(\delta_r), \quad (3.48)$$

com

$$\delta_r = \frac{\lambda_r^2}{\sigma^2}, \quad (3.49)$$

pois

$$\begin{aligned} \delta_r &= \frac{1}{\sigma^2} (\lambda_r \boldsymbol{\delta}_1^{k-r+1})^t (\lambda_r \boldsymbol{\delta}_1^{k-r+1}) = \frac{1}{\sigma^2} (\lambda_r \boldsymbol{\delta}_1^{k-r+1})^t (\lambda_r \boldsymbol{\delta}_1^{k-r+1}) = \\ &= \frac{\lambda_r^2}{\sigma^2} \end{aligned}$$

Estabeleça-se a seguinte proposição

**Proposição 3.2.3** Quando no teste (3.44),  $H_{0,r}^\circ$  não se verifica, a estatística de teste

$$\mathcal{F}_r^\circ = \frac{b_r}{k-r+1} \frac{\|\underline{\mathbf{v}}_r(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t)\|^2}{\|b\_ \text{vec}(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t; r)\|^2},$$

tem aproximadamente distribuição  $F$  não central com  $k-r+1$  e  $b_r$  graus de liberdade e parâmetro de não centralidade

$$\delta_r = \frac{\lambda_r^2}{\sigma^2},$$

escrevendo-se

$$\mathcal{F}_r^\circ \rightsquigarrow F_{k-r+1, b_r}(\delta_r).$$

**Dem.** Sabe-se por (3.22) que

$$b\_ \text{vec}(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t; r) \rightsquigarrow N(\mathbf{0}^{b_r}, \sigma^2 \mathbf{I}_{b_r}),$$

e consequentemente

$$\|b\_ \text{vec}(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t; r)\|^2 \rightsquigarrow \sigma^2 \chi_{b_r}^2.$$

Sabemos ainda por (3.48) que

$$\|\underline{\mathbf{v}}_r(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t)\|^2 \sim \sigma^2 \chi_{k-r+1}^2(\delta_r),$$

com

$$\delta_r = \frac{\lambda_r^2}{\sigma^2}.$$

Uma vez que  $b\_ \text{vec}(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t; r)$  e  $\underline{\mathbf{v}}_r(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t)$  são independentes, a estatística de teste

$$\mathcal{F}_r^\circ = \frac{\frac{\|\underline{\mathbf{v}}_r(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t)\|^2}{k-r+1}}{\frac{\|b\_ \text{vec}(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t; r)\|^2}{b_r}} = \frac{b_r}{k-r+1} \frac{\|\underline{\mathbf{v}}_r(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t)\|^2}{\|b\_ \text{vec}(\mathbf{P}^* \mathbf{S}(\mathbf{P}^*)^t; r)\|^2}$$

tem aproximadamente distribuição  $F$  não central com  $k-r+1$  e  $b_r$  graus de liberdade e parâmetro de não centralidade  $\delta_r$ , escrevendo-se

$$\mathcal{F}_r^\circ \rightsquigarrow F_{k-r+1, b_r}(\delta_r).$$

■

Observe-se ainda que em relação ao teste (3.44), pode-se ir baixando o grau até encontrar o primeiro  $\lambda_r$  que seja significativamente maior que 0. Ou seja, se para  $\lambda_r$  não rejeitar  $H_{0,r}^\circ$ , irei efectuar um novo teste (3.44) para  $\lambda_{r-1}$ . Se novamente não rejeitar  $H_{0,r-1}^\circ$ , irei efectuar uma novo teste, desta vez para  $\lambda_{r-2}$ , e assim sucessivamente, até encontrar o primeiro  $r'$  tal que para  $\lambda_{r'}$  o teste leva à rejeição de  $H_{0,r'}^\circ$ , e neste caso o modelo teria grau  $r'$ .

### 3.2.3 Inferência sobre os parâmetros de estrutura $(\beta_i)$ , $\sigma^2$ e parâmetros de não centralidade $(\delta_r)$

#### 3.2.3.1 Inferência sobre $\beta_i$

Sendo

$$\mathbf{S} = \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t \quad (3.50)$$

a decomposição espectral da matriz  $\mathbf{S}$  e  $\gamma_i$  o vector próprio da matriz  $\mathbf{S}$  associado ao valor próprio  $\theta_i$  temos

$$\mathbf{S} \gamma_i = \theta_i \gamma_i = \tilde{\beta}_i, \text{ com } i = 1, \dots, k, \quad (3.51)$$

embora só nos interessem os  $r$  primeiros pares  $(\gamma_i, \theta_i)$ .

Consideremos as matrizes ortogonais  $\mathbf{P}^*$  e  $\mathbf{P}^\circ$  tais que

$$\mathbf{P}^* = \begin{bmatrix} \gamma_1^t \\ \vdots \\ \gamma_k^t \end{bmatrix} \quad \text{e} \quad \mathbf{P}^\circ = \begin{bmatrix} \alpha_1^t \\ \vdots \\ \alpha_k^t \end{bmatrix} \quad (3.52)$$

então temos

$$\begin{aligned} \mathbf{S} (\mathbf{P}^*)^t &= \left( \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t \right) \begin{bmatrix} \gamma_1 & \cdots & \gamma_r & \cdots & \gamma_k \end{bmatrix} = \\ &= \begin{bmatrix} \left( \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t \right) \gamma_1 & \cdots & \left( \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t \right) \gamma_r & \cdots & \left( \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t \right) \gamma_k \end{bmatrix} = \\ &= \begin{bmatrix} \theta_1 \gamma_1 & \cdots & \theta_r \gamma_r & \cdots & \theta_k \gamma_k \end{bmatrix} = \begin{bmatrix} \tilde{\beta}_1 & \cdots & \tilde{\beta}_r & \cdots & \tilde{\beta}_k \end{bmatrix}. \end{aligned}$$

Admitindo que  $\gamma_i \approx \alpha_i$  e  $\theta_i \approx \lambda_i$  para  $i = 1, \dots, r$ , sendo  $\mathbf{v}_i^k(\mathbf{A})$  o  $i$ -ésimo vector linha de uma matriz  $\mathbf{A}$ , temos  $\mathbf{v}_i^k(\mathbf{P}^*) \approx \mathbf{v}_i^k(\mathbf{P}^\circ)$ , para  $i = 1, \dots, r$ , ver [Vaquinhas e Mexia, 1992], e considerando o modelo (3.9), vem

$$\begin{aligned} \mathbf{S} (\mathbf{P}^\circ)^t &= \left( \sum_{i=1}^r \lambda_i \alpha_i \alpha_i^t + \bar{\mathbf{E}} \right) \begin{bmatrix} \alpha_1 & \cdots & \alpha_r & \cdots & \alpha_k \end{bmatrix} = \\ &= \left( \sum_{i=1}^r \lambda_i \alpha_i \alpha_i^t \right) \begin{bmatrix} \alpha_1 & \cdots & \alpha_r & \cdots & \alpha_k \end{bmatrix} + \bar{\mathbf{E}} \begin{bmatrix} \alpha_1 & \cdots & \alpha_r & \cdots & \alpha_k \end{bmatrix} = \\ &= \begin{bmatrix} \left( \sum_{i=1}^r \lambda_i \alpha_i \alpha_i^t \right) \alpha_1 & \cdots & \left( \sum_{i=1}^r \lambda_i \alpha_i \alpha_i^t \right) \alpha_r & \cdots & \left( \sum_{i=1}^r \lambda_i \alpha_i \alpha_i^t \right) \alpha_k \end{bmatrix} + \bar{\mathbf{E}} (\mathbf{P}^\circ)^t = \\ &= \begin{bmatrix} \lambda_1 \alpha_1 & \cdots & \lambda_r \alpha_r & \cdots & \mathbf{0}^k \end{bmatrix} + \bar{\mathbf{E}} (\mathbf{P}^\circ)^t = \begin{bmatrix} \beta_1 & \cdots & \beta_r & \cdots & \mathbf{0}^k \end{bmatrix} + \bar{\mathbf{E}} (\mathbf{P}^\circ)^t. \end{aligned}$$



logo podemos assumir que

$$\begin{bmatrix} \tilde{\beta}_1 & \cdots & \tilde{\beta}_r & \cdots & \tilde{\beta}_k \end{bmatrix} \approx \begin{bmatrix} \beta_1 & \cdots & \beta_r & \cdots & \mathbf{0}^k \end{bmatrix} + \bar{\mathbf{E}}(\mathbf{P}^\circ)^t \quad (3.53)$$

em que

$$\tilde{\beta}_j \approx 0 \text{ com } j = r + 1, \dots, k. \quad (3.54)$$

Assim, sendo

$$\begin{aligned} \mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t &= \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_r \\ \vdots \\ \gamma_k \end{bmatrix} \underbrace{\begin{bmatrix} \tilde{\beta}_1 & \cdots & \tilde{\beta}_r & \cdots & \tilde{\beta}_k \end{bmatrix}}_{=\mathbf{S}(\mathbf{P}^*)^t} \approx \\ &\approx \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_r \\ \vdots \\ \alpha_k \end{bmatrix} \left( \begin{bmatrix} \beta_1 & \cdots & \beta_r & \cdots & \mathbf{0}^k \end{bmatrix} + \bar{\mathbf{E}}(\mathbf{P}^\circ)^t \right) = \\ &= \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_r \\ \vdots \\ \alpha_k \end{bmatrix} \begin{bmatrix} \beta_1 & \cdots & \beta_r & \cdots & \mathbf{0}^k \end{bmatrix} + \mathbf{P}^\circ \bar{\mathbf{E}}(\mathbf{P}^\circ)^t = \\ &= \mathbf{D}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) + \mathbf{P}^\circ \bar{\mathbf{E}}(\mathbf{P}^\circ)^t \approx \\ &\approx \mathbf{D}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) + \mathbf{P}^* \bar{\mathbf{E}}(\mathbf{P}^*)^t. \end{aligned}$$

Considerando

$$V = \left\| s\_ \text{vec} \left( \mathbf{P}^\circ \bar{\mathbf{E}}(\mathbf{P}^\circ)^t \right) \right\|^2, \quad (3.55)$$

podemos verificar que a informação contida em  $\mathbf{S}$  fica condensada nos seguintes pares  $(\tilde{\beta}_i, V)$ , com  $i = 1, \dots, r$ , quando o modelo (3.9) se ajusta.

Ao vector  $\tilde{\beta}_i$  dá-se o nome de  $i$ -ésimo vector de estrutura estimado e é utilizado para estimar o vector  $\beta_i = \lambda_i \alpha_i$ .

Observe-se ainda, que quando  $\gamma_i \approx \alpha_i$  e  $\theta_i \approx \lambda_i$  para  $i = 1, \dots, r$ , tem-se

$$\begin{aligned}\tilde{\beta}_i &= \theta_i \gamma_i = \mathbf{S} \gamma_i = \left( \sum_{j=1}^r \lambda_j \alpha_j \alpha_j^t + \bar{\mathbf{E}} \right) \gamma_i \approx \left( \sum_{j=1}^r \theta_j \gamma_j \gamma_j^t + \bar{\mathbf{E}} \right) \gamma_i = \\ &= \theta_i \gamma_i + \bar{\mathbf{E}} \gamma_i = \beta_i + \bar{\mathbf{E}} \gamma_i\end{aligned}$$

Assim, uma vez que  $\tilde{\beta}_i \approx \beta_i + \bar{\mathbf{E}} \gamma_i$ , e considerando a matriz de covariância de  $\tilde{\beta}_i$ ,  $\text{COV}(\tilde{\beta}_i)$ , temos

$$\text{COV}(\tilde{\beta}_i) \approx \text{COV}(\beta_i + \bar{\mathbf{E}} \gamma_i) = \text{COV}(\bar{\mathbf{E}} \gamma_i) = \gamma_i \text{COV}(\bar{\mathbf{E}}) \gamma_i^t \quad (3.56)$$

Se considerarmos o  $i$ -ésimo vector próprio  $\gamma_i$  tal que

$$\gamma_i = \begin{bmatrix} \gamma_{i,1} \\ \vdots \\ \gamma_{i,k} \end{bmatrix} \quad (3.57)$$

então por (3.56) vem

$$\begin{aligned}\text{COV}(\tilde{\beta}_i) &\approx \gamma_i \text{COV}(\bar{\mathbf{E}}) \gamma_i^t = \sigma^2 \left( \mathbf{I}_k + \sum_{j=1}^k \sum_{j'=1}^k \gamma_{i,j} \gamma_{i,j'} \right) = \\ &= \sigma^2 (\mathbf{I}_k + \gamma_i \gamma_i^t) \approx \sigma^2 \mathbf{I}_k, \text{ com } i = 1, \dots, k,\end{aligned}$$

porque sabemos que  $\|\gamma_i\| = 1$ , logo os elementos da matriz  $\gamma_i \gamma_i^t$  não serão grandes, podendo ser considerados praticamente nulos, isto é  $\gamma_i \gamma_i^t \approx \mathbf{0}$ .

Adiante, trabalharemos no caso equilibrado em que a ANOVA e técnicas associadas são robustas para a heterocedasticidade, o que nos permitirá admitir a homocedasticidade, ver [Scheffé, 1959] e [Ito, 1980].

Ora, os  $r$  primeiros vectores linha de  $\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t$  são independentes de  $\|b\_ \text{vec}(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t; r)\|^2$ . Observe-se que sendo  $\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t$  uma matriz simétrica, os seus  $r$  primeiros vectores coluna,  $\xi_1, \dots, \xi_r$ , serão idênticos aos  $r$  primeiros vectores linha e por isso também os  $r$  primeiros vectores coluna de  $\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t$  serão independentes de  $\|b\_ \text{vec}(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t; r)\|^2$ , logo os vectores de estrutura

$$\tilde{\beta}_j = (\mathbf{P}^*)^t \xi_j \text{ com } j = 1, \dots, r, \quad (3.58)$$

serão independentes de  $\|b\_ \text{vec}(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t; r)\|^2$ .

Estabeleça-se assim a proposição cuja demonstração será conforme o que se apresentou anteriormente.

**Proposição 3.2.4** *Os vectores de estrutura  $\tilde{\beta}_1, \dots, \tilde{\beta}_r$  são independentes de  $\|b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2$ .*

Como

$$\|\tilde{\beta}_j\| = \|\theta_j \gamma_j\| = \theta_j \|\gamma_j\| = \theta_j, \text{ com } j = 1, \dots, r, \quad (3.59)$$

e em geral,

$$\theta_1 \gg \theta_2 \gg \dots \quad (3.60)$$

a inferência desenvolvida para famílias estruturadas centrar-se-á nos primeiros vectores de estrutura ajustados.

Consideremos os  $\tilde{\beta}_1, \dots, \tilde{\beta}_r$  com vista a

- possíveis estudos futuros
- poder-se admitir que sendo  $\mathbf{B}_r = \|b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2$ , temos

$$\mathbf{B}_r \rightsquigarrow \sigma^2 \chi_{b_r}^2,$$

caso contrário, surgirão parâmetros de não centralidade nos denominadores dos testes  $F$ . Se tal acontecer, a potência destes testes ficaria diminuída, ver [Mexia, 1995].

### 3.2.3.2 Inferência sobre $\sigma^2$

Por (3.23) sabemos que

$$\|b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2 \rightsquigarrow \sigma^2 \chi_{b_r}^2, \quad (3.61)$$

e considerando

$$\mathbf{B}_r = \|b\_vec(\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2 \quad (3.62)$$

então um estimador aproximadamente centrado para  $\sigma^2$  será

$$\tilde{\sigma}^2 = \frac{\mathbf{B}_r}{b_r}. \quad (3.63)$$

Designemos por  $\chi_{b_r;p}^2$  o quantil de probabilidade  $p$  da distribuição  $\chi_{b_r}^2$ , ou seja

$$\Pr \left[ \frac{1}{\sigma^2} \mathbf{B}_r \leq \chi_{b_r;p}^2 \right] = p \quad (3.64)$$

teremos os diferentes intervalos de confiança de nível  $1 - q$  dados por

- $\left[ \frac{\mathbf{B}_r}{\chi^2_{b_r; 1-\frac{q}{2}}}; \frac{\mathbf{B}_r}{\chi^2_{b_r; \frac{q}{2}}} \right];$
- $\left[ \frac{\mathbf{B}_r}{\chi^2_{b_r; 1-q}}; +\infty \right];$
- $\left[ 0; \frac{\mathbf{B}_r}{\chi^2_{b_r; q}} \right];$

que permitem, por dualidade, testar a hipótese

$$H_0 : \sigma^2 = \sigma_0^2. \quad (3.65)$$

Assim, o teste de nível  $q$  :

- bilateral, rejeita  $H_0$  se o intervalo de confiança de nível  $1-q$ ,  $\left[ \frac{\mathbf{B}_r}{\chi^2_{b_r; 1-\frac{q}{2}}}; \frac{\mathbf{B}_r}{\chi^2_{b_r; \frac{q}{2}}} \right]$ , não contém  $\sigma_0^2$ ;
- unilateral esquerdo, rejeita  $H_0$  se o intervalo de confiança de nível  $1-q$ ,  $\left[ \frac{\mathbf{B}_r}{\chi^2_{b_r; 1-q}}; +\infty \right]$ , não contém  $\sigma_0^2$ ;
- unilateral direito, rejeita  $H_0$  se o intervalo de confiança de nível  $1-q$ ,  $\left[ 0; \frac{\mathbf{B}_r}{\chi^2_{b_r; q}} \right]$ , não contém  $\sigma_0^2$ .

### 3.2.3.3 Inferência sobre $\delta_r$

Para efectuarmos inferência sobre o parâmetro de não centralidade,  $\delta_r$ , vamos considerar

- estimadores centrados para este parâmetro;
- intervalos de confiança para este parâmetro;
- testes de hipóteses para este parâmetro.

O parâmetro de não centralidade,  $\delta_r$ , é um parâmetro importante pois o teste de hipóteses (3.44), referido anteriormente, pode ser reescrito da forma

$$\begin{cases} H_{0,r}^\circ : \delta_r = 0 \\ \quad \quad \quad vs \\ H_{1,r}^\circ : \delta_r > 0 \end{cases} \quad (3.66)$$

Para realizar inferência sobre  $\delta_r$ , iremos utilizar a estatística

$$\mathcal{T}_r^\circ = \frac{k-r+1}{b_r} \mathcal{F}_r^\circ, \quad (3.67)$$

que é mais manejável que a estatística  $\mathcal{F}_r^\circ - \mathcal{F}_r^\circ$  é a estatística referida na proposição 3.2.3. Assim, temos

$$\begin{aligned} \mathcal{T}_r^\circ &= \frac{k-r+1}{b_r} \frac{b_r}{k-r+1} \frac{\|\underline{\mathbf{v}}_r (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t)\|^2}{\|b_- \mathbf{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2} = \\ &= \frac{\|\underline{\mathbf{v}}_r (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t)\|^2}{\|b_- \mathbf{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2} \end{aligned}$$

Ora, a estatística  $\mathcal{T}_r^\circ$  tem a distribuição de um quociente de dois qui-quadrados, cujo numerador é um qui-quadrado não central com  $k-r+1$  graus de liberdade e parâmetro de não centralidade  $\delta_r$  e o denominador é um qui-quadrado central com  $b_r$  graus de liberdade. Sendo  $\|\underline{\mathbf{v}}_r (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t)\|^2$  e  $\|b_- \mathbf{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2$  independentes, então

$$\mathcal{T}_r^\circ \sim \bar{F}_{k-r+1, b_r}(\delta_r). \quad (3.68)$$

A função densidade da distribuição  $\bar{F}_{k-r+1, b_r}(\delta_r)$  será dada por

$$\bar{f}(z | k-r+1, b_r, \delta_r) = e^{-\frac{\delta_r}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta_r}{2}\right)^j}{j!} \bar{f}(z | k-r+1+2j, b_r) \quad (3.69)$$

onde, com  $h = k-r+1+2j$ , se tem a função densidade da distribuição  $\bar{F}_{h, b_r}$  dada por

$$\bar{f}(z | h, b_r) = \begin{cases} 0 & , \quad z < 0 \\ \frac{\Gamma\left(\frac{h+b_r}{2}\right)}{\Gamma\left(\frac{h}{2}\right)\Gamma\left(\frac{b_r}{2}\right)} \frac{z^{\frac{h}{2}-1}}{(1+z)^{\frac{h+b_r}{2}}} & , \quad z > 0 \end{cases} \quad (3.70)$$

onde  $\Gamma(\cdot)$  é a função gama com  $\Gamma(p) = \int_0^{+\infty} u^{p-1} e^{-u} du$ , com  $p \in \mathbb{R}^+$ , ver [Mexia, 1995].

Sabemos que integrando por partes, obtém-se a seguinte propriedade para a função gama

$$\Gamma(p+1) = p\Gamma(p) \quad (3.71)$$

e que

$$\int_{-\infty}^{+\infty} \bar{f}(z | h, b_r) dz = 1 \quad (3.72)$$

por se tratar de uma função densidade de probabilidade.

Assim, ver [Oliveira, 2001], temos

$$\int_0^{+\infty} \frac{z^{\frac{h}{2}-1}}{(1+z)^{\frac{h+b_r}{2}}} dz = \frac{\Gamma\left(\frac{h}{2}\right) \Gamma\left(\frac{b_r}{2}\right)}{\Gamma\left(\frac{h+b_r}{2}\right)}. \quad (3.73)$$

Sendo

$$\mathcal{T}_r^\circ \sim \bar{F}_{k-r+1, b_r}(\delta_r), \quad (3.74)$$

e com  $b_r > 2$ , temos o valor médio dado por

$$\mu(z | k-r+1, b_r, \delta_r) = e^{-\frac{\delta_r}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta_r}{2}\right)^j}{j!} \mu(z | k-r+1+2j, b_r), \quad (3.75)$$

em que, considerando  $h = k-r+1+2j$ , se tem

$$\mu(\cdot | h, b_r) = \frac{h}{b_r - 2}. \quad (3.76)$$

Logo

$$\mu(z | k-r+1, b_r, \delta_r) = \frac{1}{b_r - 2} (k-r+1 + \delta_r), \quad (3.77)$$

o que faz com que

$$\tilde{\delta}_r = (b_r - 2) \mathcal{T}_r^\circ - (k-r+1). \quad (3.78)$$

seja um estimador centrado para  $\delta_r$ .

Por outro lado, a distribuição  $\bar{F}_{k-r+1, b_r}(\delta_r)$  é, ver [Mexia, 1995], estritamente decrescente em  $\delta_r$ , pelo que podemos utilizar o procedimento geral para obter intervalos de confiança para  $\delta_r$ , ver [Mood *et al.*, 1987].

Começemos por obter uma estatística que depende de  $\delta_r$ , mas cuja distribuição não depende de  $\delta_r$ . Então, essa estatística vai ser

$$\mathcal{T}_r^\circ = \frac{k-r+1}{b_r} \mathcal{F}_r^\circ \quad (3.79)$$

tal que

$$\mathcal{T}_r^\circ \sim \bar{F}_{k-r+1, b_r}(\delta_r). \quad (3.80)$$

Assim, usando a distribuição de  $\mathcal{T}_r^\circ$ , e resolvendo em ordem a  $\delta_r$  a equação

$$\bar{F}(\mathcal{T}_r^\circ | k-r+1, b_r, \delta_r) = 1-q \quad (3.81)$$

obtém-se o limite inferior para um intervalo de confiança de nível  $1-q$ , limitado à esquerda. Na prática, é este valor que nos interessa, pois se tivermos "confiança" de que  $\delta_r > 0$  somos levados a aceitar  $r$  como limite inferior do grau.

No entanto, observe-se que para valores grandes de  $\delta_r$  verificou-se que a solução numérica da equação (3.81) levanta algumas dificuldades. Com vista a solucionar estas dificuldades, iremos recorrer à linearidade assintótica apresentada na secção seguinte (secção 3.2.3.4).

### 3.2.3.4 Uma aplicação da lineariedade assintótica

Para resolvermos a dificuldade apresentada na secção anterior vamos recorrer à lineariedade assintótica, ver [Mexia e Oliveira, 2010], [Mexia *et al.*, 2011] e [Nunes *et al.*, 2013].

Esta técnica aplica-se, com vista a obtenção de distribuições limite ligadas ao crescimento da não centralidade e não da dimensão da amostra.

Essas distribuições limite permitir-nos-ão abordar questões relativas à predominância em situações, em que o primeiro valor próprio ou os primeiros valores próprios de uma matriz simétrica são fortemente dominantes.

Comecemos por apresentar a definição de uma função assintoticamente linear.

**Definição 3.2.1 (*Função Assintoticamente Linear*)** Dada uma função suficientemente regular  $g : \mathbb{R}^k \rightarrow \mathbb{R}$ , sejam  $\underline{g}$  e  $\underline{\underline{g}}$  o gradiente<sup>1</sup> e a matriz hessiana<sup>2</sup> de  $g$ , respectivamente. Então, sendo  $r_d(\mathbf{x})$  o supremo do raio espectral  $r(\mathbf{y})$  de  $\underline{\underline{g}}(\mathbf{y})$ <sup>3</sup>, quando  $\|\mathbf{y} - \mathbf{x}\| \leq d$ , a função  $g$  é uma **função assintoticamente linear**, se qualquer que seja  $d > 0$ , temos

$$k_d(u) \xrightarrow{u \rightarrow +\infty} 0$$

com

$$k_d(u) = \sup \left\{ \frac{r_d(\mathbf{x})}{\|\underline{g}(\mathbf{x})\|}; \|\mathbf{x}\| \geq u \right\}.$$

---


$$\begin{aligned} {}^1 \underline{g}(\mathbf{x}) &= \left[ \frac{\partial g}{\partial x_1} \frac{\partial g}{\partial x_2} \cdots \frac{\partial g}{\partial x_k} \right]^t \\ {}^2 \underline{\underline{g}}(\mathbf{x}) &= \begin{bmatrix} \frac{\partial^2 g}{\partial x_1^2} & \cdots & \frac{\partial^2 g}{\partial x_1 \partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial x_k \partial x_1} & \cdots & \frac{\partial^2 g}{\partial x_k^2} \end{bmatrix} \end{aligned}$$

<sup>3</sup>Designamos por raio espectral de uma matriz  $\mathbf{A}$  o valor :

$$\rho(\mathbf{A}) = \max_{i=1, \dots, n} |\lambda_i(\mathbf{A})|$$

onde  $\lambda_i(\mathbf{A})$  são os valores próprios de  $\mathbf{A}$ .

**Observação:** Se a matriz  $\mathbf{A}$  for simétrica então  $\|\mathbf{A}\| = \rho(\mathbf{A})$ , pois  $\|\mathbf{A}\| = \sqrt{\rho(\mathbf{A}^T \mathbf{A})}$ .

Seja  $\mathbf{A}$  uma matriz quadrada. Então para qualquer norma matricial, temos  $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$ ; e para qualquer  $c > 0$ , existe sempre uma norma tal que  $\|\mathbf{A}\| \leq \rho(\mathbf{A}) + c$ , ou seja, o raio espectral é o ínfimo do conjunto das normas de uma matriz.

Vê-se que uma dada função  $g$  é uma função assintoticamente linear se ao afastarmos-nos da origem, a sua componente linear tornar-se dominante. Por exemplo, sabendo que as primeiras e as segundas derivadas parciais de polinómios são elas mesmas polinómios com grau decrescente com derivações sucessivas, então os polinómios serão assintoticamente lineares.

Note-se que formas lineares e quadráticas, bem como outros polinómios de baixo grau desempenham um papel importante na inferência estatística. Assim, atendendo à importância de polinómios de baixo grau e com o objectivo de estudar as estatísticas que sejam da forma

$$\mathbf{Y} = g(\mathbf{X} + \boldsymbol{\mu}), \quad (3.82)$$

consideremos  $g$  uma função assintoticamente linear, sendo  $\mathbf{X}$  um vector aleatório  $k$ -dimensional, com uma distribuição qualquer e  $\boldsymbol{\mu} \in \mathbb{R}^k$  um vector não aleatório tal que

$$\|\boldsymbol{\mu}\| \rightarrow +\infty. \quad (3.83)$$

Consideremos

$$\mathbf{Z} = g(\boldsymbol{\mu}) + \underline{g}(\boldsymbol{\mu})^t \mathbf{X} \quad (3.84)$$

o que é equivalente a

$$\mathbf{Z} - g(\boldsymbol{\mu}) = \underline{g}(\boldsymbol{\mu})^t \mathbf{X}. \quad (3.85)$$

Consideremos também

$$\mathbf{Z}^\circ = \frac{\mathbf{Z} - g(\boldsymbol{\mu})}{\|\underline{g}(\boldsymbol{\mu})\|} = \frac{\underline{g}(\boldsymbol{\mu})^t \mathbf{X}}{\|\underline{g}(\boldsymbol{\mu})\|} \quad (3.86)$$

e

$$\mathbf{Y}^\circ = \frac{\mathbf{Y} - g(\boldsymbol{\mu})}{\|\underline{g}(\boldsymbol{\mu})\|} = \frac{g(\mathbf{X} + \boldsymbol{\mu}) - g(\boldsymbol{\mu})}{\|\underline{g}(\boldsymbol{\mu})\|}. \quad (3.87)$$

Provou-se, ver [Mexia e Oliveira, 2010] e [Mexia *et al.*, 2011], que as funções de distribuição  $F_{\mathbf{Z}}$  e  $F_{\mathbf{Z}^\circ}$  de  $\mathbf{Z}$  e  $\mathbf{Z}^\circ$ , respectivamente, são boas aproximações das funções de distribuição  $F_{\mathbf{Y}}$  e  $F_{\mathbf{Y}^\circ}$ , funções de distribuição de  $\mathbf{Y}$  e  $\mathbf{Y}^\circ$ , respectivamente.

Mostrou-se também, ver [Mexia e Oliveira, 2010] e [Mexia *et al.*, 2011], que

$$\sup \{ |F_{\mathbf{Y}^\circ}(\mathbf{z}) - F_{\mathbf{Z}^\circ}(\mathbf{z})| \} \xrightarrow{\|\boldsymbol{\mu}\| \rightarrow +\infty} 0, \quad (3.88)$$

sendo  $F_{\mathbf{Y}^\circ}$  e  $F_{\mathbf{Z}^\circ}$  as funções de distribuição de  $\mathbf{Y}^\circ$  e  $\mathbf{Z}^\circ$ , respectivamente.

Observe-se, que se a distribuição  $F_{\mathbf{Z}^\circ}$  não depender de  $\boldsymbol{\mu}$  (ou de  $\frac{g(\boldsymbol{\mu})}{\|\underline{g}(\boldsymbol{\mu})\|}$ ), tem-se

$$F_{\mathbf{Z}^\circ}(\mathbf{z}) = F^\circ(\mathbf{z}) \quad (3.89)$$



e nesse caso teremos

$$\sup \{|F_{\mathbf{Y}^\circ}(\mathbf{z}) - F^\circ(\mathbf{z})|\} \xrightarrow{\|\boldsymbol{\mu}\| \rightarrow +\infty} 0, \quad (3.90)$$

e

$$F_{\mathbf{Y}^\circ} \xrightarrow[\|\boldsymbol{\mu}\| \rightarrow +\infty]{u} F^\circ \quad (3.91)$$

considerando, em geral, que  $F_{\mathbf{Y}^\circ}(\mathbf{z})$  não depende de  $\boldsymbol{\mu}$  e onde  $\xrightarrow{u}$  representa a convergência uniforme.

Uma vez que as distribuições qui-quadrado e  $t$  de *Student* possuem numerosas aplicações importantes em inferência estatística e desempenham um papel importante, em particular nesta dissertação, segue um exemplo de uma aplicação a estas distribuições, com vista a obtenção de intervalos de confiança para o parâmetro de não centralidade.

Consideremos para o efeito, a função  $g$  tal que

$$g(\mathbf{X}) = \|\mathbf{X}\|^2. \quad (3.92)$$

Temos então

$$\underline{g}(\mathbf{X}) = \left( \frac{\partial g}{\partial X_1}, \frac{\partial g}{\partial X_2}, \dots, \frac{\partial g}{\partial X_k} \right) = (2X_1, \dots, 2X_k) = 2(X_1, \dots, X_k) = 2\mathbf{X}$$

e

$$\begin{aligned} \underline{\underline{g}}(x) &= \begin{bmatrix} \frac{\partial^2 g}{\partial X_1^2} & \cdots & \frac{\partial^2 g}{\partial X_1 \partial X_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial X_k \partial X_1} & \cdots & \frac{\partial^2 g}{\partial X_k^2} \end{bmatrix} = \begin{bmatrix} 2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 2 \end{bmatrix} = \\ &= 2 \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} = 2\mathbf{I}_k. \end{aligned}$$

pelo que  $g$  é uma função assintoticamente linear.

Tomemos

$$\mathbf{Y} = g(\mathbf{X} + \boldsymbol{\mu}) = \|\mathbf{X} + \boldsymbol{\mu}\|^2 \quad (3.93)$$

com  $\mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_k)$ . Assim teremos

$$\mathbf{Y} \sim \sigma^2 \chi_k^2(\delta), \quad (3.94)$$

ou seja,  $\mathbf{Y}$  tem distribuição do produto de  $\sigma^2$  por um qui-quadrado com  $k$  graus de liberdade e parâmetro de não centralidade  $\delta$ , em que

$$\delta = \frac{\|\boldsymbol{\mu}\|^2}{\sigma^2} = \frac{g(\boldsymbol{\mu})}{\sigma^2}. \quad (3.95)$$

Como

$$\underline{g}(\boldsymbol{\mu}) = 2\boldsymbol{\mu} \quad (3.96)$$

temos

$$\|\underline{g}(\boldsymbol{\mu})\| = \|2\boldsymbol{\mu}\| = 2\|\boldsymbol{\mu}\| = 2\sigma\sqrt{\delta},$$

pois por (3.95) temos

$$\|\boldsymbol{\mu}\| = \sigma\sqrt{\delta}. \quad (3.97)$$

Observe-se agora que por (3.87), temos, com  $U = \|\mathbf{X} + \boldsymbol{\mu}\|^2$

$$\begin{aligned} \mathbf{Y}^\circ &= \frac{\mathbf{Y} - g(\boldsymbol{\mu})}{\|\underline{g}(\boldsymbol{\mu})\|} = \frac{g(\mathbf{X} + \boldsymbol{\mu}) - g(\boldsymbol{\mu})}{\|\underline{g}(\boldsymbol{\mu})\|} = \\ &= \frac{U - \|\boldsymbol{\mu}\|^2}{2\sigma\sqrt{\delta}} = \frac{U - \sigma^2\delta}{2\sigma\sqrt{\delta}} \end{aligned}$$

logo tendo em conta que

$$\begin{aligned} \frac{\mathbf{Y}^\circ}{\sigma} &= \frac{\frac{U - \sigma^2\delta}{2\sigma\sqrt{\delta}}}{\sigma} = \frac{U - \sigma^2\delta}{2\sigma^2\sqrt{\delta}} = \\ &= \frac{\frac{U}{\sigma^2} - \delta}{2\sqrt{\delta}} \hookrightarrow N(0, 1) \end{aligned}$$

pois  $\mathbf{Y}^\circ \sim N(0, \sigma^2)$  e onde  $\hookrightarrow$  indica distribuição limite.

Assim, sendo  $\mathbf{Y}^\circ \sim N(0, \sigma^2)$  independente de  $S \sim \sigma^2\chi_g^2$ , temos

$$L = \frac{\mathbf{Y}^\circ}{\sqrt{\frac{S}{g}}} \sim t_g, \quad (3.98)$$

isto é,  $L$  segue uma distribuição  $t$  de *Student* com  $g$  graus de liberdade. Portanto

$$\frac{\frac{U - \sigma^2\delta}{2\sigma\sqrt{\delta}}}{\sqrt{\frac{S}{g}}} \sim t_g \quad (3.99)$$

e sendo  $u$  o valor de  $U$  e  $s$  o valor de  $S$ , faz com que,

$$\begin{aligned} u - \sigma^2\delta &= 2\sigma\sqrt{\delta}\sqrt{\frac{s}{g}}t_g \Leftrightarrow u - \sigma^2\delta - 2\sigma\sqrt{\delta}\sqrt{\frac{s}{g}}t_g = 0 \Leftrightarrow \\ &\Leftrightarrow \sigma^2\delta + 2\sigma\sqrt{\delta}\sqrt{\frac{s}{g}}t_g - u = 0 \end{aligned}$$

e resolvendo em ordem a  $\sqrt{\delta}$  tem-se

$$\begin{aligned}\sqrt{\delta} &= \frac{-\sigma\sqrt{\frac{s}{g}}t_g + \sqrt{\sigma^2\frac{s}{g}t_g^2 + \sigma^2u}}{\sigma^2} \Leftrightarrow \\ \Leftrightarrow \sqrt{\delta} &= \frac{1}{\sigma} \left( \sqrt{\frac{s}{g}}t_g + \sqrt{\frac{s}{g}t_g^2 + u} \right)\end{aligned}$$

Uma vez que  $S \sim \sigma^2\chi_g^2$  então  $\frac{S}{\chi_g^2}$  será um bom estimador para  $\sigma^2$ , logo uma estimativa para  $\frac{1}{\sigma}$  será  $\sqrt{\frac{\chi_g^2}{s}}$ .

Teremos agora que gerar pares  $(t_{g(i)}, \chi_{g(i)}^2)$  com  $t_{g(i)} \sim t_g$  e  $\chi_{g(i)}^2 \sim \chi_g^2$  tomando para estimador de  $\sqrt{\delta}$

$$\widetilde{\sqrt{\delta}}_{(i)} = \sqrt{\frac{\chi_{g(i)}^2}{s}} \left( \sqrt{\frac{s}{g}}t_{g(i)} + \sqrt{\frac{s}{g}t_{g(i)}^2 + u} \right), \quad (3.100)$$

o que nos permite efectuar simulações.

Ordenando as estimativas obtidas para  $\sqrt{\delta}$ , temos

$$\sqrt{\delta}_{(1)} < \sqrt{\delta}_{(2)} < \dots < \sqrt{\delta}_{(N)} \quad (3.101)$$

e considerando  $\widetilde{\sqrt{\delta}}_p$  o quantil de ordem  $p$  este será dado por

$$\widetilde{\sqrt{\delta}}_p = \sqrt{\delta}_{(pN)} \quad (3.102)$$

e portanto o intervalo de confiança a  $(1 - q) \times 100\%$  de confiança para  $\sqrt{\delta}$  será

$$\left[ \widetilde{\sqrt{\delta}}_{\frac{q}{2}}; \widetilde{\sqrt{\delta}}_{1-\frac{q}{2}} \right]. \quad (3.103)$$

O teorema de Glivenko-Cantelli, ver [Loève, 1977], diz-nos que à medida que a dimensão da amostra aumenta, a distribuição empírica tende para a distribuição teórica, ao comparar os quantis empíricos e teóricos, ou seja, se considerarmos  $F_n$  a função de distribuição empírica de uma amostra de dimensão  $n$  constituída por observações *i.i.d.* com distribuição  $F$  então tem-se

$$S_n = \sup \{|F_n(x) - F(x)|\} \xrightarrow[n \rightarrow +\infty]{q.c.} 0 \quad (3.104)$$

onde  $\xrightarrow{q.c.}$  significa convergência quase certa.

Em [Ferreira, 2005], apresenta-se um resultado análogo para a função inversa da função de distribuição. Sendo  $F(x)$  uma função contínua e estritamente crescente, qualquer que seja  $p \in ]0, 1]$ , o quantil de  $F$  para a probabilidade  $p$  será dado por

$$x_p = F^{-1}(p). \quad (3.105)$$

Seja ainda  $x_{n,p}$  o quantil de  $F_n$  para a probabilidade  $p$ . Assim

**Proposição 3.2.5** *Se  $F(x)$  tiver função densidade de probabilidade contínua  $f(x)$ , e se  $f(x) > 0$  sempre que  $0 < F(x) < 1$ , qualquer que seja  $\alpha \in ]0, 1[$  tem-se*

$$S_{n,\alpha} = \sup \left\{ |x_{n,p} - x_p| : \frac{\alpha}{2} < p < 1 - \frac{\alpha}{2} \right\} \xrightarrow[n \rightarrow +\infty]{q.c.} 0$$

Assim, dispondo dos pares  $(t_{g(i)}, \chi_{g(i)}^2)$  com  $i = 1, \dots, N$  em que  $t_{g(i)}$  segue uma distribuição  $t$  com  $g(i)$  graus de liberdade e é independente de  $\chi_{g(i)}^2$  que segue uma distribuição qui-quadrado central com  $g(i)$  graus de liberdade, podem-se obter os quantis empíricos para  $\sqrt{\delta}$  e a partir daqui construir intervalos de confiança para  $\sqrt{\delta}$  e consequentemente para  $\delta$ .

No exemplo de aplicação prática apresentado em 5.3.1, não foi necessário, por razões que adiante apresentamos, aplicar estes resultados.

# Capítulo 4

## Famílias Estruturadas

### 4.1 Considerações Prévias

Neste capítulo vamos considerar famílias estruturadas de matrizes estocásticas simétricas. As matrizes de uma tal família terão a mesma dimensão, sendo no nosso caso  $k \times k$ , e corresponderão aos  $n$  tratamentos de um delineamento base.

Ao estudar-se uma tal família procura-se analisar a acção dos factores do delineamento base sobre os modelos da família considerada.

Os delineamentos que se consideram são delineamentos ortogonais estando associados a partições ortogonais de  $\mathbb{R}^n$ . Assim, como veremos, podemos aplicar a ANOVA e técnicas associadas na análise.

Admitindo-se que os modelos da família têm graus  $r_1, \dots, r_n$  não nulos, e considerando-se

$$r = \min \{r_1, \dots, r_n\} \quad (4.1)$$

com  $r > 0$ , pode-se estudar a acção dos factores do delineamento base sobre os  $r$  primeiros vectores de estrutura.

No entanto, como referimos atrás, muitas vezes o primeiro valor próprio é fortemente dominante para a generalidade dos modelos, tendo-se  $r = 1$ . Em tais situações temos de analisar a acção dos factores sobre o primeiro vector de estrutura.

Quando o grau é o mesmo para todas as matrizes da família, as mesmas contribuem com o mesmo número de graus de liberdade para o erro. Estamos assim no caso equilibrado e poderá admitir-se a homocedasticidade, ver [Scheffé, 1959] e [Ito, 1980].

## 4.2 Delineamento Base - Estrutura Algébrica

Como referimos, os delineamentos base que estamos a considerar têm estrutura ortogonal, estando associados a partições da forma

$$\mathbb{R}^n = \boxplus_{j=1}^m \nabla_j \quad (4.2)$$

onde  $\boxplus$  indica soma directa ortogonal.

Dado um vector de observações com vector médio  $\boldsymbol{\mu}$  interessará testar hipóteses da forma

$$H_{0,j}(\mathbf{b}_j) : \boldsymbol{\mu} - \mathbf{b}_j \in \nabla_j^\perp, j = 1, \dots, m, \quad (4.3)$$

onde  $\perp$  indica o complemento ortogonal.

Estas hipóteses generalizam as hipóteses usuais da forma

$$H_{0,j} : \boldsymbol{\mu} \in \nabla_j^\perp, j = 1, \dots, m, \quad (4.4)$$

tratando-se de uma generalização que faz surgir, como veremos, propriedades da dualidade para os testes  $F$ .

Isto sugere que a família das hipóteses  $H_{0,j}(\mathbf{b}_j), j = 1, \dots, m$  constitui uma formulação natural das hipóteses a testar.

Estes delineamentos são, como veremos, fáceis de tratar, quando se dispõe de um par  $(\mathbf{Y}, S)$  em que

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{C}) \quad (4.5)$$

com  $\mathbf{C}$  uma matriz conhecida e definida positiva, independente de

$$S \sim \sigma^2 \chi_g^2. \quad (4.6)$$

Na secção 4.3 mostraremos como construir testes  $F$  e obter resultados relacionados para cada uma das hipóteses

$$H_{0,j}(\mathbf{b}_j), j = 1, \dots, m. \quad (4.7)$$

## 4.3 Heterocedasticidade controlada sem restrições

Vejamos agora, como construir a estatística de teste para testes de hipóteses da forma

$$H_0 : \mathbf{A}\boldsymbol{\mu} = \mathbf{A}\mathbf{b}. \quad (4.8)$$

em que  $\mathbf{A}$  é uma matriz do tipo  $k \times n$ .

Suponhamos que se tem um par aleatório  $(\mathbf{Y}, S)$ , em que

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{C}) \quad (4.9)$$

e

$$S \sim \sigma^2 \chi_g^2, \quad (4.10)$$

sendo  $\mathbf{Y}$  independente de  $S$  e  $\mathbf{C}$  uma matriz conhecida e definida positiva. Como a matriz de covariância de  $\mathbf{Y}$  é uma matriz conhecida, a menos de  $\sigma^2$ , diremos que temos a **heterocedasticidade controlada** [Mexia, 1989a]. Diremos ainda que **não se tem restrições** pois pode admitir-se que  $\boldsymbol{\mu}$  é um vector qualquer de  $\mathbb{R}^n$ , sendo  $n$  o número de componentes de  $\mathbf{Y}$ , e pode-se utilizar  $S$  para estimar  $\sigma^2$ .

Se considerarmos

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{0}\}, \quad (4.11)$$

o espaço de nulidade de  $\mathbf{A}$ , a hipótese  $H_0$  em (4.8) pode ser reescrita da forma

$$H_0 : \boldsymbol{\mu} - \mathbf{b} \in \mathcal{N}(\mathbf{A}), \quad (4.12)$$

pois

$$\mathbf{A}\boldsymbol{\mu} = \mathbf{A}\mathbf{b} \Leftrightarrow \mathbf{A}\boldsymbol{\mu} - \mathbf{A}\mathbf{b} = \mathbf{0}^k \Leftrightarrow \mathbf{A}(\boldsymbol{\mu} - \mathbf{b}) = \mathbf{0}^k. \quad (4.13)$$

Por outro lado, considerando

$$\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\mu} \quad (4.14)$$

e

$$\boldsymbol{\psi}_0 = \mathbf{A}\mathbf{b}, \quad (4.15)$$

temos a proposição 4.3.1 que nos permite reescrever a hipótese  $H_0$  em (4.8) da forma

**Proposição 4.3.1** *A hipótese  $H_0 : \mathbf{A}\boldsymbol{\mu} = \mathbf{A}\mathbf{b}$  pode ser reescrita como*

$$H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0.$$

**Dem.** Quando  $\boldsymbol{\psi} = \boldsymbol{\psi}_0$  temos  $\mathbf{A}\boldsymbol{\mu} = \mathbf{A}\mathbf{b}$  vindo  $\boldsymbol{\mu} - \mathbf{b} \in \mathcal{N}(\mathbf{A})$ . Inversamente, se  $\boldsymbol{\mu} - \mathbf{b} \in \mathcal{N}(\mathbf{A})$  então  $\mathbf{A}(\boldsymbol{\mu} - \mathbf{b}) = \mathbf{0}^k$ , logo  $\mathbf{A}\boldsymbol{\mu} - \mathbf{A}\mathbf{b} = \mathbf{0}^k$ , que é equivalente a  $\mathbf{A}\boldsymbol{\mu} = \mathbf{A}\mathbf{b}$ , o que é o mesmo que ter-se  $\boldsymbol{\psi} = \boldsymbol{\psi}_0$ . ■

Por outro lado se escrevermos

$$\mathbf{u} \rho_{\mathcal{N}(\mathbf{A})} \mathbf{v} \quad (4.16)$$

quando e só quando  $\mathbf{A}\mathbf{u} = \mathbf{A}\mathbf{v}$ , e definimos assim uma relação de congruência  $\rho_{\mathcal{N}(\mathbf{A})}$  em  $\mathbb{R}^n$ .

Assim sendo, representamos por  $[\mathbf{u}]_{\mathcal{N}(\mathbf{A})}$  a classe de congruência a que  $\mathbf{u}$  pertence. Vê-se assim, que a hipótese  $H_0$  em (4.8) pode, uma vez mais, ser reescrita de outra forma, ou seja

$$H_0 : \boldsymbol{\mu} \in [\mathbf{b}]_{\mathcal{N}(\mathbf{A})}, \quad (4.17)$$

pois  $\boldsymbol{\mu} \in [\mathbf{b}]_{\mathcal{N}(\mathbf{A})}$  significa que  $\boldsymbol{\mu} \rho_{\mathcal{N}(\mathbf{A})} \mathbf{b}$ , ou seja,  $\mathbf{A}\boldsymbol{\mu} = \mathbf{A}\mathbf{b}$ .

Por outro lado, considerando  $\tilde{\boldsymbol{\psi}} = \mathbf{A}\mathbf{Y}$ , temos

$$\tilde{\boldsymbol{\psi}} \sim N(\boldsymbol{\psi}, \sigma^2 \mathbf{A}\mathbf{C}\mathbf{A}^t), \quad (4.18)$$

pois

$$E[\tilde{\boldsymbol{\psi}}] = E[\mathbf{A}\mathbf{Y}] = \mathbf{A}E[\mathbf{Y}] = \mathbf{A}\boldsymbol{\mu} = \boldsymbol{\psi} \quad (4.19)$$

e que

$$V[\tilde{\boldsymbol{\psi}}] = V[\mathbf{A}\mathbf{Y}] = \mathbf{A}V[\mathbf{Y}]\mathbf{A}^t = \mathbf{A}(\sigma^2 \mathbf{C})\mathbf{A}^t = \sigma^2 \mathbf{A}\mathbf{C}\mathbf{A}^t. \quad (4.20)$$

Interessa-nos agora estabelecer o seguinte lema

**Lema 4.3.1** *Quando os vectores linha da matriz  $\mathbf{A}$  são linearmente independentes e  $\mathbf{C}$  é uma matriz definida positiva<sup>1</sup>,  $\mathbf{A}\mathbf{C}\mathbf{A}^t$  também é uma matriz definida positiva.*

**Dem.** Como os vectores linha da matriz  $\mathbf{A}$  são linearmente independentes, os vectores coluna da matriz  $\mathbf{A}^t$  também são linearmente independentes.

Consideremos  $\mathbf{z} \neq \mathbf{0}^k$ , vindo  $\mathbf{A}^t\mathbf{z} \neq \mathbf{0}^n$ , e como  $\mathbf{C}$  é uma matriz definida positiva, temos

$$\mathbf{z}^t (\mathbf{A}\mathbf{C}\mathbf{A}^t) \mathbf{z} = (\mathbf{A}^t\mathbf{z})^t \mathbf{C} (\mathbf{A}^t\mathbf{z}) > 0$$

o que implica que  $\mathbf{z}^t (\mathbf{A}\mathbf{C}\mathbf{A}^t) \mathbf{z} > 0$ , o que estabelece a tese de  $\mathbf{A}\mathbf{C}\mathbf{A}^t$  ser uma matriz definida positiva, pois visto  $\mathbf{z}$  ser um vector qualquer não nulo. ■

Uma vez que se admite que a matriz  $\mathbf{C}$  é uma matriz definida positiva, recorde-se que, matrizes de covariância regulares são matrizes definidas positivas<sup>2</sup>, ver [Mexia, 1989a]. Observe-se que  $\sigma^2 > 0$ , e sendo  $\sigma^2 \mathbf{C}$  uma matriz definida positiva, então é equivalente a  $\sigma^2 \mathbf{C}$  ser uma matriz invertível. Assim, neste caso diremos que temos **heterocedasticidade controlada regular**.

Assim  $\mathbf{A}\mathbf{C}\mathbf{A}^t$  é uma matriz invertível, logo uma matriz definida positiva, se e só se  $\mathbf{A}$  for horizontalmente livre, isto é, os vectores linha da matriz  $\mathbf{A}$  são linearmente independentes.

Portanto, interessa-nos estabelecer a seguinte proposição

---

<sup>1</sup> $\mathbf{C}$  é uma matriz definida positiva se  $\mathbf{x}^t \mathbf{C} \mathbf{x} > 0$ , para todo o  $\mathbf{x} \in \mathbb{R}^n$ .

<sup>2</sup>Sabemos que quando uma matriz é definida positiva, significa que os seus valores próprios são todos positivos e consequentemente o determinante da matriz é diferente de 0, e por isso a matriz é invertível, logo a matriz é regular.



**Proposição 4.3.2** *Sendo a heterocedasticidade controlada regular e sendo  $\mathbf{A}$  uma matriz horizontalmente livre do tipo  $k \times n$ , a estatística*

$$\mathcal{F} = \frac{g}{k} \frac{\left(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right)^t (\mathbf{A}\mathbf{C}\mathbf{A}^t)^{-1} \left(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right)}{S}$$

*tem distribuição  $F$  não central com  $k$  e  $g$  graus de liberdade e parâmetro de não centralidade*

$$\delta = \frac{1}{\sigma^2} (\boldsymbol{\psi} - \boldsymbol{\psi}_0)^t (\mathbf{A}\mathbf{C}\mathbf{A}^t)^{-1} (\boldsymbol{\psi} - \boldsymbol{\psi}_0)$$

*e escrevendo-se*

$$\mathcal{F} \sim F_{k,g}(\delta).$$

**Dem.** Considerando

$$\tilde{\boldsymbol{\psi}} = \mathbf{A}\mathbf{Y} \quad , \quad \boldsymbol{\psi} = \mathbf{A}\boldsymbol{\mu} \quad \text{e} \quad \boldsymbol{\psi}_0 = \mathbf{A}\mathbf{b},$$

temos

$$\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0 \sim N(\boldsymbol{\psi} - \boldsymbol{\psi}_0, \sigma^2 \mathbf{A}\mathbf{C}\mathbf{A}^t)$$

pois

$$\begin{aligned} E[\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0] &= E[\mathbf{A}\mathbf{Y} - \mathbf{A}\mathbf{b}] = E[\mathbf{A}\mathbf{Y}] - E[\mathbf{A}\mathbf{b}] = \\ &= \mathbf{A}E[\mathbf{Y}] - \mathbf{A}\mathbf{b} = \mathbf{A}\boldsymbol{\mu} - \mathbf{A}\mathbf{b} = \boldsymbol{\psi} - \boldsymbol{\psi}_0 \end{aligned}$$

e

$$\begin{aligned} V[\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0] &= V[\mathbf{A}\mathbf{Y} - \mathbf{A}\mathbf{b}] = V[\mathbf{A}\mathbf{Y}] = \mathbf{A}V[\mathbf{Y}]\mathbf{A}^t = \\ &= \mathbf{A}(\sigma^2 \mathbf{C})\mathbf{A}^t = \sigma^2 \mathbf{A}\mathbf{C}\mathbf{A}^t. \end{aligned}$$

Temos também

$$\left(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right)^t (\mathbf{A}\mathbf{C}\mathbf{A}^t)^{-1} \left(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right) \sim \sigma^2 \chi_k^2(\delta)$$

pois sendo

$$\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0 \sim N(\boldsymbol{\psi} - \boldsymbol{\psi}_0, \sigma^2 \mathbf{A}\mathbf{C}\mathbf{A}^t),$$

a forma quadrática é o produto por  $\sigma^2$  de um qui-quadrado com  $k$  graus de liberdade e parâmetro de não centralidade  $\delta$ , ver [Mexia, 1990], em que

$$\delta = \frac{1}{\sigma^2} (\boldsymbol{\psi} - \boldsymbol{\psi}_0)^t (\mathbf{A}\mathbf{C}\mathbf{A}^t)^{-1} (\boldsymbol{\psi} - \boldsymbol{\psi}_0).$$

Uma vez que  $\mathbf{Y}$  e  $S$  são independentes temos também que  $(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$  e  $S$  também são independentes, e consequentemente, a forma quadrática é independente de

$S$ , e daí resulta directamente a tese, já que, o quociente entre um qui-quadrado não central a dividir pelos seus graus de liberdade e um qui-quadrado central a dividir pelos seus graus de liberdade, resulta numa distribuição  $F$  não central com  $k$  e  $g$  graus de liberdade e parâmetro de não centralidade  $\delta$ , isto é,

$$\mathcal{F} = \frac{\frac{(\tilde{\psi} - \psi_0)^t (\mathbf{A} \mathbf{C} \mathbf{A}^t)^{-1} (\tilde{\psi} - \psi_0)}{\frac{k}{\frac{S}{g}}}}{\frac{g}{k}} = \frac{g}{k} \frac{(\tilde{\psi} - \psi_0)^t (\mathbf{A} \mathbf{C} \mathbf{A}^t)^{-1} (\tilde{\psi} - \psi_0)}{S}$$

tal que

$$\mathcal{F} \sim F_{k,g}(\delta)$$

em que o parâmetro de não centralidade é igual a

$$\delta = \frac{1}{\sigma^2} (\psi - \psi_0)^t (\mathbf{A} \mathbf{C} \mathbf{A}^t)^{-1} (\psi - \psi_0).$$

■

Considerando-se agora que a matriz  $\mathbf{A}$ , pode ser vista como

$$\mathbf{A}^t = [\mathbf{a}_1 \dots \mathbf{a}_k] \quad (4.21)$$

em que  $\mathbf{a}_1 \dots \mathbf{a}_k$  são vectores coluna formados pelas linhas da matriz  $\mathbf{A}$ , poderemos considerar também as seguintes sub-hipóteses

$$H_{0,i} : \mathbf{a}_i^t \boldsymbol{\mu} = \mathbf{a}_i^t \mathbf{b}, \text{ com } i = 1, \dots, k. \quad (4.22)$$

Note-se que estas hipóteses derivam da hipótese  $H_0$  em (4.8) e portanto verifica-se a hipótese  $H_0$  se e só se se verificarem as hipóteses  $H_{0,i}$ , com  $i = 1, \dots, k$ , o que nos leva a dizer que a hipótese  $H_0$  é a intersecção das hipóteses  $H_{0,i}$ , com  $i = 1, \dots, k$ , escrevendo-se

$$H_0 = \bigcap_{i=1}^k H_{0,i}. \quad (4.23)$$

Por outro lado, diremos que as componentes de  $\tilde{\psi} = \mathbf{A} \mathbf{Y}$ ,  $\psi = \mathbf{A} \boldsymbol{\mu}$  e  $\psi_0 = \mathbf{A} \mathbf{b}$  são as seguintes

$$\tilde{\psi}_i = \mathbf{a}_i^t \mathbf{Y}, \text{ com } i = 1, \dots, k; \quad (4.24)$$

$$\psi_i = \mathbf{a}_i^t \boldsymbol{\mu}, \text{ com } i = 1, \dots, k; \quad (4.25)$$

$$\psi_{0,i} = \mathbf{a}_i^t \mathbf{b}, \text{ com } i = 1, \dots, k, \quad (4.26)$$

respectivamente.

Assim temos

$$\tilde{\psi}_i \sim N(\psi_i, \sigma^2 \mathbf{a}_i^t \mathbf{C} \mathbf{a}_i), \text{ com } i = 1, \dots, k, \quad (4.27)$$

e

$$\tilde{\psi}_i - \psi_{0,i} \sim N(\psi_i - \psi_{0,i}, \sigma^2 \mathbf{a}_i^t \mathbf{C} \mathbf{a}_i), \text{ com } i = 1, \dots, k, \quad (4.28)$$

pois

$$E[\tilde{\psi}_i] = E[\mathbf{a}_i^t \mathbf{Y}] = \mathbf{a}_i^t E[\mathbf{Y}] = \mathbf{a}_i^t \boldsymbol{\mu} = \psi_i,$$

$$\begin{aligned} E[\tilde{\psi}_i - \psi_{0,i}] &= E[\mathbf{a}_i^t \mathbf{Y} - \mathbf{a}_i^t \mathbf{b}] = E[\mathbf{a}_i^t \mathbf{Y}] - E[\mathbf{a}_i^t \mathbf{b}] = \mathbf{a}_i^t E[\mathbf{Y}] - \mathbf{a}_i^t \mathbf{b} = \\ &= \mathbf{a}_i^t \boldsymbol{\mu} - \mathbf{a}_i^t \mathbf{b} = \psi_i - \psi_{0,i}, \end{aligned}$$

$$V[\tilde{\psi}_i] = V[\mathbf{a}_i^t \mathbf{Y}] = \mathbf{a}_i^t V[\mathbf{Y}] \mathbf{a}_i = \mathbf{a}_i^t (\sigma^2 \mathbf{C}) \mathbf{a}_i = \sigma^2 \mathbf{a}_i^t \mathbf{C} \mathbf{a}_i$$

e

$$\begin{aligned} V[\tilde{\psi}_i - \psi_{0,i}] &= V[\mathbf{a}_i^t \mathbf{Y} - \mathbf{a}_i^t \mathbf{b}] = V[\mathbf{a}_i^t \mathbf{Y}] = \mathbf{a}_i^t V[\mathbf{Y}] \mathbf{a}_i = \\ &= \mathbf{a}_i^t (\sigma^2 \mathbf{C}) \mathbf{a}_i = \sigma^2 \mathbf{a}_i^t \mathbf{C} \mathbf{a}_i. \end{aligned}$$

Observe-se que  $\tilde{\psi}_i$  é independente de  $S$ , bem como  $\tilde{\psi}_i - \psi_{0,i}$  também é independente de  $S$ . Assim, para testarmos a hipótese  $H_{0,i}$ , temos a seguinte estatística de teste

$$\mathcal{F}_i = \frac{\frac{1}{\mathbf{a}_i^t \mathbf{C} \mathbf{a}_i} (\tilde{\psi}_i - \psi_{0,i})^2}{\frac{1}{g} S} = \frac{g}{\mathbf{a}_i^t \mathbf{C} \mathbf{a}_i} \frac{(\tilde{\psi}_i - \psi_{0,i})^2}{S}, \text{ com } i = 1, \dots, k \quad (4.29)$$

que tem distribuição  $F$  não central com 1 e  $g$  graus de liberdade e parâmetro de não centralidade

$$\delta_i = \frac{(\psi_i - \psi_{0,i})^2}{\sigma^2 \mathbf{a}_i^t \mathbf{C} \mathbf{a}_i}, \text{ com } i = 1, \dots, k. \quad (4.30)$$

Por outro lado, temos a seguinte proposição

**Proposição 4.3.3** *As variáveis pivot*

$$\mathcal{F}' = \frac{g}{k} \frac{(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi})^t (\mathbf{A} \mathbf{C} \mathbf{A}^t)^{-1} (\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi})}{S}$$

e

$$\mathcal{F}'_i = \frac{g}{\mathbf{a}_i^t \mathbf{C} \mathbf{a}_i} \frac{(\tilde{\psi}_i - \psi_i)^2}{S}, \text{ com } i = 1, \dots, k$$

têm distribuições  $F$  centrais com  $k$  e  $g$  graus de liberdade e com 1 e  $g$  graus de liberdade, respectivamente.

**Dem.** Considerando  $\tilde{\boldsymbol{\psi}} = \mathbf{A}\mathbf{Y}$ ,  $\boldsymbol{\psi} = \mathbf{A}\boldsymbol{\mu}$  e  $\boldsymbol{\psi}_0 = \mathbf{A}\mathbf{b}$ , temos

$$\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi} \sim N(\mathbf{0}^k, \sigma^2 \mathbf{A}\mathbf{C}\mathbf{A}^t),$$

pois

$$\begin{aligned} E[\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}] &= E[\mathbf{A}\mathbf{Y} - \mathbf{A}\boldsymbol{\mu}] = E[\mathbf{A}\mathbf{Y}] - E[\mathbf{A}\boldsymbol{\mu}] = \mathbf{A}E[\mathbf{Y}] - \mathbf{A}\boldsymbol{\mu} = \\ &= \mathbf{A}\boldsymbol{\mu} - \mathbf{A}\boldsymbol{\mu} = \mathbf{0}^k \end{aligned}$$

e

$$V[\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}] = V[\mathbf{A}\mathbf{Y} - \mathbf{A}\boldsymbol{\mu}] = V[\mathbf{A}\mathbf{Y}] = \mathbf{A}V[\mathbf{Y}]\mathbf{A}^t = \mathbf{A}(\sigma^2\mathbf{C})\mathbf{A}^t = \sigma^2\mathbf{A}\mathbf{C}\mathbf{A}^t$$

e consequentemente temos

$$(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi})^t (\mathbf{A}\mathbf{C}\mathbf{A}^t)^{-1} (\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}) \sim \sigma^2 \chi_k^2.$$

Temos também que

$$\tilde{\psi}_i - \psi_i \sim N(0, \sigma^2 \mathbf{a}_i^t \mathbf{C} \mathbf{a}_i),$$

pois

$$\begin{aligned} E[\tilde{\psi}_i - \psi_i] &= E[\mathbf{a}_i^t \mathbf{Y} - \mathbf{a}_i^t \boldsymbol{\mu}] = E[\mathbf{a}_i^t \mathbf{Y}] - E[\mathbf{a}_i^t \boldsymbol{\mu}] = \mathbf{a}_i^t E[\mathbf{Y}] - \mathbf{a}_i^t \boldsymbol{\mu} = \\ &= \mathbf{a}_i^t \boldsymbol{\mu} - \mathbf{a}_i^t \boldsymbol{\mu} = 0 \end{aligned}$$

e

$$V[\tilde{\psi}_i - \psi_i] = V[\mathbf{a}_i^t \mathbf{Y} - \mathbf{a}_i^t \boldsymbol{\mu}] = V[\mathbf{a}_i^t \mathbf{Y}] = \mathbf{a}_i^t V[\mathbf{Y}] \mathbf{a}_i = \mathbf{a}_i^t (\sigma^2 \mathbf{C}) \mathbf{a}_i = \sigma^2 \mathbf{a}_i^t \mathbf{C} \mathbf{a}_i$$

e consequentemente temos

$$\frac{(\tilde{\psi}_i - \psi_i)^2}{\mathbf{a}_i^t \mathbf{C} \mathbf{a}_i} \sim \sigma^2 \chi_1^2.$$

Assim, uma vez que

$$S \sim \sigma^2 \chi_g^2,$$

temos

$$\mathcal{F}' = \frac{\frac{(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi})^t (\mathbf{A}\mathbf{C}\mathbf{A}^t)^{-1} (\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi})}{k}}{\frac{S}{g}} = \frac{g}{k} \frac{(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi})^t (\mathbf{A}\mathbf{C}\mathbf{A}^t)^{-1} (\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi})}{S}$$

em que  $\mathcal{F}'$  tem distribuição  $F$  central com  $k$  e  $g$  graus de liberdade, e escreve-se

$$\mathcal{F}' \sim F_{k,g}$$

e temos também

$$\mathcal{F}'_i = \frac{\frac{(\tilde{\psi}_i - \psi_i)^2}{\mathbf{a}_i^t \mathbf{C} \mathbf{a}_i}}{\frac{1}{\frac{S}{g}}} = \frac{\frac{(\tilde{\psi}_i - \psi_i)^2}{\mathbf{a}_i^t \mathbf{C} \mathbf{a}_i}}{\frac{S}{g}} = \frac{g}{\mathbf{a}_i^t \mathbf{C} \mathbf{a}_i} \frac{(\tilde{\psi}_i - \psi_i)^2}{S},$$

em que  $\mathcal{F}'_i$  tem distribuição  $F$  central com 1 e  $g$  graus de liberdade, e escreve-se

$$\mathcal{F}'_i \sim F_{1,g}$$

com  $i = 1, \dots, k$ . ■

Da variável pivot  $\mathcal{F}'$  obtém-se elipsóides de confiança para  $\boldsymbol{\psi}$ , já que considerando  $f_{p,k,g}$  o quantil de ordem  $p$  para a distribuição  $F$  central com  $k$  e  $g$  graus de liberdade, tem-se que

$$\begin{aligned} & \Pr \left( \left( \boldsymbol{\psi} - \tilde{\boldsymbol{\psi}} \right)^t (\mathbf{A} \mathbf{C} \mathbf{A}^t)^{-1} \left( \boldsymbol{\psi} - \tilde{\boldsymbol{\psi}} \right) \leq k \times f_{1-q,k,g} \frac{S}{g} \right) = \\ &= \Pr \left( \frac{g}{k} \frac{\left( \boldsymbol{\psi} - \tilde{\boldsymbol{\psi}} \right)^t (\mathbf{A} \mathbf{C} \mathbf{A}^t)^{-1} \left( \boldsymbol{\psi} - \tilde{\boldsymbol{\psi}} \right)}{S} \leq f_{1-q,k,g} \right) = \\ &= \Pr (\mathcal{F}' \leq f_{1-q,k,g}) = 1 - q. \end{aligned}$$

Observe-se que o teste de nível  $q$  rejeita  $H_0$ , se e só se,  $\boldsymbol{\psi}_0$  não estiver contido no elipsóide de confiança de nível  $1 - q$ . Assim o teste F goza de dualidade.

Por outro lado, um ponto está no interior de um elipsóide se e só se estiver entre todos os pares de planos paralelos tangentes ao elipsóide.

Dado  $\mathbf{a} \neq \mathbf{0}$  existe um tal par de planos tangentes e ortogonais a  $\mathbf{a}$ , estando  $\boldsymbol{\psi}$  entre os seus pares (ver [Scheffé, 1959]) se e só se

$$\left| \mathbf{a}^t \boldsymbol{\psi} - \mathbf{a}^t \tilde{\boldsymbol{\psi}} \right| \leq \sqrt{k \times f_{1-q,k,g} \mathbf{a}^t (\mathbf{A} \mathbf{C} \mathbf{A}^t) \mathbf{a} \frac{S}{g}} \quad (4.31)$$

pelo que temos os intervalos de confiança simultâneos dados por

$$\Pr \left( \bigcap_{\mathbf{a}} \left| \mathbf{a}^t \boldsymbol{\psi} - \mathbf{a}^t \tilde{\boldsymbol{\psi}} \right| \leq \sqrt{k \times f_{1-q,k,g} \mathbf{a}^t (\mathbf{A} \mathbf{C} \mathbf{A}^t) \mathbf{a} \frac{S}{g}} \right) = 1 - q, \quad (4.32)$$

onde  $\bigcap_{\mathbf{a}}$  indica que se consideram todos os vectores  $\mathbf{a}$  com  $k$  componentes.<sup>3</sup>

---

<sup>3</sup>Observe-se que para  $\mathbf{a} = \mathbf{0}$ , se tem a desigualdade trivial  $0 \leq 0$ .

Esta última expressão pode ser considerada como uma versão do bem conhecido teorema de Scheffé, ver [Mexia, 1989b].

Observe-se que quando

$$\sqrt{k \times f_{1-q,k,g} \mathbf{a}^t (\mathbf{A} \mathbf{C} \mathbf{A}^t) \mathbf{a} \frac{S}{g}} \leq |\mathbf{a}^t \mathbf{b} - \mathbf{a}^t \boldsymbol{\psi}| \quad (4.33)$$

se tem

$$\Pr (\mathbf{a}^t \boldsymbol{\psi} = \mathbf{a}^t \mathbf{b}) \leq q \quad (4.34)$$

dizendo-se que  $\mathbf{a}^t \boldsymbol{\psi}$  é significativamente diferente de  $\mathbf{a}^t \mathbf{b}$  ao nível de probabilidade  $q$ .

Quando se utiliza o teorema de Scheffé para encontrar diferenças significativas entre as componentes de  $\boldsymbol{\psi}$ , trabalha-se com os vectores da forma

$$\boldsymbol{\delta}_i - \boldsymbol{\delta}_{i'} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ -1 \\ \vdots \\ 0 \end{bmatrix} \quad (4.35)$$

onde 1, corresponde à  $i$ -ésima componente do vector  $\boldsymbol{\delta}_i - \boldsymbol{\delta}_{i'}$  e  $-1$  à  $i'$ -ésima componente do vector  $\boldsymbol{\delta}_i - \boldsymbol{\delta}_{i'}$ , e com  $\mathbf{b} = \mathbf{0}$ , então  $\psi_i$  e  $\psi_{i'}$  diferem significativamente ao nível  $q$  se

$$\sqrt{k f_{1-q,k,g} (u_{ii} - 2u_{ii'} + u_{i'i'}) \frac{S}{g}} < |\tilde{\psi}_i - \tilde{\psi}_{i'}| \quad (4.36)$$

já que com  $\mathbf{A} \mathbf{C} \mathbf{A}^t = \mathbf{U} = [u_{ij}]$  com  $i = 1, \dots, k$  e  $j = 1, \dots, k$ , se tem

$$(\boldsymbol{\delta}_i - \boldsymbol{\delta}_{i'})^t \mathbf{U} (\boldsymbol{\delta}_i - \boldsymbol{\delta}_{i'}) = u_{ii} - 2u_{ii'} + u_{i'i'}. \quad (4.37)$$

Por outro lado, para as restantes variáveis pivot,  $\mathcal{F}'_i$ , com  $i = 1, \dots, k$ , obtém-se os intervalos de confiança dados pelos pares de desigualdades

$$\tilde{\psi}_i - \sqrt{\mathbf{a}_i^t \mathbf{C} \mathbf{a}_i f_{1-q,1,g} \frac{S}{g}} \leq \psi_i \leq \tilde{\psi}_i + \sqrt{\mathbf{a}_i^t \mathbf{C} \mathbf{a}_i f_{1-q,1,g} \frac{S}{g}}, \quad (4.38)$$

com  $i = 1, \dots, k$ , já que estas são equivalentes às desigualdades

$$\left( \tilde{\psi}_i - \psi_i \right)^2 \leq \mathbf{a}_i^t \mathbf{C} \mathbf{a}_i f_{1-q,1,g} \frac{S}{g}, \text{ com } i = 1, \dots, k, \quad (4.39)$$

e

$$\Pr \left( \left( \tilde{\psi}_i - \psi_i \right)^2 \leq \mathbf{a}_i^t \mathbf{C} \mathbf{a}_i f_{1-q,1,g} \frac{S}{g} \right) = \Pr \left( \mathcal{F}'_i \leq f_{1-q,1,g} \right) = 1 - q, \quad (4.40)$$

com  $i = 1, \dots, k$ .

Observe-se que os testes  $F$ , (4.8) e (4.22), que construímos gozam de dualidade

- o teste de nível  $q$  para  $H_0$  não rejeita essa hipótese se e só se  $\psi_o$  estiver contido no elipsóide de confiança de nível  $1 - q$ ;
- o teste de nível  $q$  para  $H_{0,i}$ , com  $i = 1, \dots, k$ , não rejeita essa hipótese se e só se  $\psi_{o,i}$ , com  $i = 1, \dots, k$  estiver contido no intervalo de confiança de nível  $1 - q$ .

Por outro lado as hipóteses testadas,  $H_0$  e  $H_{0,i}$ , podem também ser reescritas da forma

$$H_0 : \delta = 0 \quad (4.41)$$

e

$$H_{0,i} : \delta_i = 0, \text{ com } i = 1, \dots, k, \quad (4.42)$$

respectivamente, pelo que os testes  $F$  são estritamente não distorcidos.

Assim, sendo  $F_{k,g}(\delta)$  a distribuição  $F$  não central com  $k$  e  $g$  graus de liberdade e parâmetro de não centralidade  $\delta$ , tem-se que, [Mexia, 1989a],

$$\frac{\partial F(z|k, g, \delta)}{\partial \delta} < 0 \quad (4.43)$$

pelo que a potência do teste de nível  $q$  é dada por

$$pot_q(\delta) = 1 - F(f_{1-q,k,g}|k, g, \delta) \quad (4.44)$$

e que vai crescendo com  $\delta$ .

Analogamente, sendo  $F_{k,g}(\delta_i)$  a distribuição  $F$  não central com 1 e  $g$  graus de liberdade e parâmetro de não centralidade  $\delta_i$ , tem-se que

$$\frac{\partial F(z|1, g, \delta_i)}{\partial \delta_i} < 0, \text{ com } i = 1, \dots, k, \quad (4.45)$$

logo os testes para as hipóteses  $H_{0,i}$ , com  $i = 1, \dots, k$  também são estritamente não distorcidos.

Observe-se que estamos a admitir que  $\mathbf{A} \mathbf{C} \mathbf{A}^t$  é uma matriz definida positiva.

Quando se tem um delineamento base ortogonal em vez de uma hipótese  $H_0(\mathbf{b})$  haverá que considerar as hipóteses  $H_{0,j}(\mathbf{b})$ ,  $j = 1, \dots, m$ . Podem assim,

ser aplicados os resultados desta secção desde que se disponha de matrizes  $\mathbf{A}_j$ , com  $j = 1, \dots, m$ , cujos vectores linha constituem uma base ortonormada para  $\nabla_j$ , com  $j = 1, \dots, m$ , representando-se

$$\mathbf{A}_j \longleftrightarrow \nabla_j, \text{ com } j = 1, \dots, m \quad (4.46)$$

quando queremos indicar que os vectores linha da matriz  $\mathbf{A}_j$  constituem uma base ortonormada para  $\nabla_j$ , com  $j = 1, \dots, m$ . Então

$$\mathbf{P} = [\mathbf{A}_1^t \dots \mathbf{A}_m^t]^t \quad (4.47)$$

será uma matriz ortogonal associada à partição.

Consideremos  $\mathbf{b} \in \mathbb{R}^n$  e sendo  $\mathbf{P}$  é uma matriz ortogonal, considere-se a família ortogonal de hipóteses da forma

$$H_0(\mathbf{b}, j) : \mathbf{A}_j \boldsymbol{\mu} = \mathbf{A}_j \mathbf{b}, \text{ com } j = 1, \dots, m \quad (4.48)$$

sendo

$$\mathbf{A}_j^t = [\mathbf{a}_{j,1} \dots \mathbf{a}_{j,k_j}], \text{ com } j = 1, \dots, m. \quad (4.49)$$

Podemos ainda considerar as seguintes sub-hipóteses

$$H_{0,i}(\mathbf{b}, j) : \mathbf{a}_{j,i}^t \boldsymbol{\mu} = \mathbf{a}_{j,i}^t \mathbf{b}, \text{ com } i = 1, \dots, k_j \text{ e } j = 1, \dots, m. \quad (4.50)$$

Sendo

$$\tilde{\boldsymbol{\psi}}_j = \mathbf{A}_j \mathbf{Y}, \text{ com } j = 1, \dots, m; \quad (4.51)$$

$$\boldsymbol{\psi}_j = \mathbf{A}_j \boldsymbol{\mu}, \text{ com } j = 1, \dots, m; \quad (4.52)$$

e

$$\boldsymbol{\psi}_{0,j} = \mathbf{A}_j \mathbf{b}, \text{ com } j = 1, \dots, m \quad (4.53)$$

tem-se para as hipóteses  $H_0(\mathbf{b}, j)$ , com  $j = 1, \dots, m$ , referidas em (4.48), as seguintes estatísticas de teste

$$\mathcal{F}(\mathbf{b}, j) = \frac{g}{k_j} \frac{(\tilde{\boldsymbol{\psi}}_j - \boldsymbol{\psi}_{0,j})^t (\mathbf{A}_j \mathbf{C} \mathbf{A}_j^t)^{-1} (\tilde{\boldsymbol{\psi}}_j - \boldsymbol{\psi}_{0,j})}{S}, \text{ com } j = 1, \dots, m, \quad (4.54)$$

que seguem uma distribuição  $F$  não central com  $k_j$  e  $g$  graus de liberdade e parâmetros de não centralidade

$$\delta(\mathbf{b}, j) = \frac{1}{\sigma^2} (\boldsymbol{\psi}_j - \boldsymbol{\psi}_{0,j})^t (\mathbf{A}_j \mathbf{C} \mathbf{A}_j^t)^{-1} (\boldsymbol{\psi}_j - \boldsymbol{\psi}_{0,j}), \text{ com } j = 1, \dots, m. \quad (4.55)$$

Note-se que os parâmetros de não centralidade só se anulam quando as hipóteses  $H_0(\mathbf{b}, j)$  correspondentes são verdadeiras. Assim estes testes serão estritamente não distorcidos.



Considere-se

$$\tilde{\psi}_{j,1}, \dots, \tilde{\psi}_{j,k_j} \quad (4.56)$$

as componentes de  $\tilde{\psi}_j$ , com  $j = 1, \dots, m$ ; e

$$\psi_{j,1}, \dots, \psi_{j,k_j} \quad (4.57)$$

as componentes de  $\psi_j$ , com  $j = 1, \dots, m$ ; bem como

$$\psi_{j,0,1}, \dots, \psi_{j,0,k_j} \quad (4.58)$$

as componentes de  $\psi_{j,0}$ , com  $j = 1, \dots, m$ .

Para as sub-hipóteses  $H_{0,i}(\mathbf{b}, j)$ , com  $i = 1, \dots, k_j$  e  $j = 1, \dots, m$ , temos as seguintes estatísticas de teste

$$\mathcal{F}_i(\mathbf{b}, j) = \frac{g}{v_{j,i}} \frac{\left(\tilde{\psi}_{j,i} - \psi_{j,0,i}\right)^2}{S}, \text{ com } i = 1, \dots, k_j \text{ e } j = 1, \dots, m, \quad (4.59)$$

que têm distribuição  $F$  não central com 1 e  $g$  graus de liberdade e parâmetros de não centralidade

$$\delta_i(\mathbf{b}, j) = \frac{1}{v_{j,i}\sigma^2} (\psi_{j,i} - \psi_{j,0,i}), \text{ com } i = 1, \dots, k_j \text{ e } j = 1, \dots, m, \quad (4.60)$$

sendo

$$v_{j,i} = \mathbf{a}_{j,i}^t \mathbf{C} \mathbf{a}_{j,i}, \text{ com } i = 1, \dots, k_j \text{ e } j = 1, \dots, m. \quad (4.61)$$

Estes parâmetros de não centralidade,  $\delta_i(\mathbf{b}, j)$ , anulam-se quando e só quando as correspondentes sub-hipóteses  $H_{0,i}(\mathbf{b}, j)$  se verificam, pelo que também estes testes são estritamente não distorcidos.

**Proposição 4.3.4** *Para  $\psi_j$ , com  $j = 1, \dots, m$  temos as seguintes variáveis pivot*

$$\mathcal{F}'(\mathbf{b}, j) = \frac{g}{k_j} \frac{\left(\tilde{\psi}_j - \psi_j\right)^t (\mathbf{A}_j \mathbf{C} \mathbf{A}_j^t)^{-1} \left(\tilde{\psi}_j - \psi_j\right)}{S}, \text{ com } j = 1, \dots, m,$$

*que seguem uma distribuição  $F$  central com  $k_j$  e  $g$  graus de liberdade, com  $j = 1, \dots, m$ .*

Raciocinando de forma análoga ao apresentado anteriormente, obtêm-se para os elipsóides de confiança de nível  $1 - q$

$$\left(\psi_j - \tilde{\psi}_j\right)^t (\mathbf{A}_j \mathbf{C} \mathbf{A}_j^t)^{-1} \left(\psi_j - \tilde{\psi}_j\right) \leq k_j \times f_{1-q, k_j, g} \frac{S}{g}, \quad (4.62)$$

com  $j = 1, \dots, m$ , enquanto que o teorema de Scheffé dá os intervalos de confiança simultâneos expressos por

$$\Pr \left( \bigcap_{\mathbf{a}_j} \left( \left| \mathbf{a}_j^t \psi_j - \mathbf{a}_j^t \tilde{\psi}_j \right| \leq \sqrt{k_j \times f_{1-q, k_j, g}} \mathbf{a}_j^t (\mathbf{A}_j \mathbf{C} \mathbf{A}_j^t) \mathbf{a}_j \frac{S}{g} \right) \right) = 1 - q, \quad (4.63)$$

com  $j = 1, \dots, m$ .

Mais uma vez, estes testes gozam da dualidade já que o teste de nível de significância  $q$  não rejeita  $H_0(\mathbf{b}, j)$  quando e só quando  $\psi_j$  está contido no elipsóide de confiança de nível  $1 - q$ , com  $j = 1, \dots, m$ .

**Proposição 4.3.5** *Para os  $\psi_{j,i}$ , com  $i = 1, \dots, k_j$  e  $j = 1, \dots, m$ , temos as seguintes variáveis pivot*

$$\mathcal{F}'_i(\mathbf{b}, j) = \frac{g}{v_{j,i}} \frac{(\tilde{\psi}_{j,i} - \psi_{j,i})^2}{S}, \quad \text{com } i = 1, \dots, k_j \text{ e } j = 1, \dots, m,$$

que têm distribuição  $F$  central com 1 e  $g$  graus de liberdade.

Neste caso, tem-se os intervalos de confiança de nível  $1 - q$ , dados por

$$\tilde{\psi}_{j,i} - \sqrt{v_{j,i} \times f_{1-q, 1, g} \frac{S}{g}} \leq \psi_{j,i} \leq \tilde{\psi}_{j,i} + \sqrt{v_{j,i} \times f_{1-q, 1, g} \frac{S}{g}}, \quad (4.64)$$

com  $i = 1, \dots, k_j$  e  $j = 1, \dots, m$ , mostrando-se ainda que também estes testes para as sub-hipóteses gozam de dualidade.

### 4.3.1 Análise conjunta de matrizes estocásticas simétricas

Começemos por considerar os primeiros vectores de estrutura ajustados

$$\tilde{\beta}_1(1), \dots, \tilde{\beta}_1(m) \quad (4.65)$$

para as  $m$  matrizes de uma família estruturada, a que correspondem as somas de quadrados

$$S_i \sim \sigma^2 \chi_{g_i}^2, \quad \text{com } i = 1, \dots, m. \quad (4.66)$$

Podemos admitir a independência das matrizes, pelo que os pares  $(\tilde{\beta}_1(i), S_i)$ , com  $i = 1, \dots, m$ , serão independentes. Atendendo à reprodutibilidade dos quadrados temos

$$S = \sum_{i=1}^m S_i \sim \sigma^2 \chi_g^2 \quad (4.67)$$

com

$$g = \sum_{i=1}^m g_i \quad (4.68)$$

independente de qualquer vector  $\mathbf{Y}$  com componentes

$$Y_i = \mathbf{c}^t \tilde{\boldsymbol{\beta}}_1(i), \text{ com } i = 1, \dots, m \quad (4.69)$$

o qual terá matriz de covariância  $\sigma^2 \mathbf{C}$  com

$$\mathbf{C} = \mathbf{D} \left( \mathbf{c}^t (\mathbf{I}_k + \gamma_1(1) \gamma_1^t(1)) \mathbf{c}, \dots, \mathbf{c}^t (\mathbf{I}_k + \gamma_1(m) \gamma_1^t(m)) \mathbf{c} \right), \quad (4.70)$$

uma vez que a matriz de covariância de  $\tilde{\boldsymbol{\beta}}_1(i)$  será  $\mathbf{I}_k + \gamma_1(i) \gamma_1^t(i)$ , para  $i = 1, \dots, m$ , porque sabemos que  $\|\gamma_1(i)\| = 1$ , logo os elementos da matriz  $\gamma_1(i) \gamma_1^t(i)$  não serão grandes, podendo ser considerados praticamente nulos, isto é  $\gamma_1(i) \gamma_1^t(i) \approx \mathbf{0}$ . Por isso, muitas vezes pode-se admitir a homocedasticidade aproximada dos vectores de estrutura ajustados,  $\tilde{\boldsymbol{\beta}}_1(1), \dots, \tilde{\boldsymbol{\beta}}_1(m)$ , uma vez que  $\gamma_1(i) \gamma_1^t(i) \approx \mathbf{0}$ , tendo-se nesse caso  $\mathbf{C} = \mathbf{I}_n$ , ver [Oliveira, 2001].

Por outro lado quando se querem comparar as  $j$ -ésimas componentes dos primeiros vectores de estrutura toma-se o vector

$$\mathbf{c} = \boldsymbol{\delta}_j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}, j = 1, \dots, k, \quad (4.71)$$

onde 1 corresponde à  $j$ -ésima componente de  $\boldsymbol{\delta}_j$  e neste caso dizemos que a análise é transversal.

Se as várias componentes desses vectores correspondem a momentos sucessivos, poderá tomar-se  $\mathbf{c}$  como um vector de contrastes. Por exemplo, podemos tomar

$$\mathbf{c} = \frac{1}{\sqrt{2}} (\boldsymbol{\delta}_k - \boldsymbol{\delta}_1) = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \vdots \\ 0 \\ 0 \\ \vdots \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad (4.72)$$

onde este vector de contraste poderá ser utilizado para medir a evolução global e neste caso estaríamos em presença de uma análise longitudinal.

Os resultados sobre heterocedasticidade controlada poderão agora ser aplicáveis aos pares  $(\mathbf{Y}, S)$ .

### 4.3.2 Álgebras de Jordan Comutativas e Cruzamento de Factores

As partições ortogonais que consideramos na secção 4.2, traduzem a existência de um modelo com estrutura ortogonal.

Um caso muito importante é quando os tratamentos correspondem a todas as combinações possíveis dos níveis de  $u$  factores, dizendo-se nesse caso que os factores cruzam.

As Álgebras de Jordan Comutativas de matrizes simétricas, são muito úteis para exprimir a estrutura destes modelos. Na secção 2.2 introduzimos as Álgebras de Jordan Comutativas, abordando alguns dos aspectos apresentados nesta secção.

Consideremos  $u$  factores de efeitos fixos, com  $a_1, \dots, a_u$  níveis para cada efeito e quando cruzamos os factores, temos

$$\dot{n} = \prod_{i=1}^u a_i \quad (4.73)$$

combinações possíveis dos níveis, sendo  $\dot{n}$  o número total de tratamentos, ver [Khuri *et al.*, 1998].

Num modelo deste tipo tem de se considerar os efeitos dos factores e as interacções entre esses mesmos factores. Esses efeitos e essas interacções corresponderão aos vectores não nulos de

$$\Gamma = \{\mathbf{h} : h_l = 0 \vee h_l = 1, \text{ com } l = 1, \dots, u\}. \quad (4.74)$$

A  $\mathbf{h} = \mathbf{0}$  podemos fazer corresponder o valor médio geral, enquanto que se  $\mathbf{h}$  tiver uma componente não nula corresponder-lhe-ão os efeitos do factor indicado por essa componente. Caso  $\mathbf{h}$  tenha mais do que uma componente não nula corresponder-lhe-ão as interacções entre os factores indicados por essas componentes.

Podemos ordenar os vectores segundos os índices

$$i(\mathbf{h}) = 1 + \sum_{j=1}^u h_j 2^{j-1}. \quad (4.75)$$

Os modelos de efeitos fixos para o conjunto de  $u$  factores podem, ver [Fonseca *et al.*, 2006], no caso equilibrado, ser escritos como

$$\mathbf{Y} = \sum_{\mathbf{h} \in \Gamma} \mathbf{X}(\mathbf{h}) \boldsymbol{\eta}(\mathbf{h}) + \mathbf{e} \quad (4.76)$$

onde, se se tiver  $r$  observações por tratamento, se tem

$$\mathbf{X}(\mathbf{h}) = \bigotimes_{l=1}^u \mathbf{X}_l(h_l) \otimes \mathbf{1}^r, \text{ com } \mathbf{h} \in \Gamma \quad (4.77)$$

onde  $\bigotimes_{l=1}^u$  representa o produto de Kronecker entre  $u$  matrizes e consideram-se as matrizes

$$\begin{cases} \mathbf{X}_l(0) = \mathbf{1}^{a_l}, & \text{com } l = 1, \dots, u \\ \mathbf{X}_l(1) = \mathbf{I}_{a_l}, & \text{com } l = 1, \dots, u \end{cases} \quad (4.78)$$

tendo-se ainda

$$\mathbf{e} \sim N(\mathbf{0}^n, \sigma^2 \mathbf{I}_n) \quad (4.79)$$

com  $n = \dot{n}r$ .

Nesta reformulação, o vector médio não será um vector qualquer de  $\mathbb{R}^n$ , mas sim um vector do espaço imagem de matrizes  $\mathbf{I}_{\dot{n}} \otimes \mathbf{1}^r$ .

Teremos pois de adaptar a formulação apresentada ao modelo que estamos a considerar. Assim, consideremos o modelo

$$\mathbf{Y} = \sum_{\mathbf{h} \in \Gamma} \mathbf{X}(\mathbf{h}) \boldsymbol{\eta}(\mathbf{h}) + \mathbf{e} \quad (4.80)$$

em que se tem  $u$  factores de efeitos fixos com  $a_1, \dots, a_u$  níveis para cada efeito, tomando-se

$$\mathbf{X}(\mathbf{h}) = \bigotimes_{l=1}^u \mathbf{X}_l(h_l) \quad (4.81)$$

e em que  $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{C})$ .

Assim, temos

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{C}) \quad (4.82)$$

com

$$\boldsymbol{\mu} = \sum_{\mathbf{h} \in \Gamma} \mathbf{X}(\mathbf{h}) \boldsymbol{\eta}(\mathbf{h}) \quad (4.83)$$

e que é independente de

$$S \sim \sigma^2 \chi_g^2. \quad (4.84)$$

Sejam

$$\mathbf{M}(\mathbf{h}) = \mathbf{X}(\mathbf{h}) (\mathbf{X}(\mathbf{h}))^t, \text{ com } \mathbf{h} \in \Gamma, \quad (4.85)$$

que são matrizes que também se escrevem da forma

$$\mathbf{M}(\mathbf{h}) = \bigotimes_{l=1}^u \mathbf{M}_l(h_l), \text{ com } \mathbf{h} \in \Gamma, \quad (4.86)$$

e em que

$$\begin{cases} \mathbf{M}_l(0) = \mathbf{J}_{a_l}, & \text{com } l = 1, \dots, u \\ \mathbf{M}_l(1) = \mathbf{I}_{a_l}, & \text{com } l = 1, \dots, u \end{cases} \quad (4.87)$$

onde  $\mathbf{J}_s = \mathbf{1}^s (\mathbf{1}^s)^t$  é a matriz do tipo  $s \times s$  com todos os elementos iguais a 1.

Ora se considerarmos

$$\mathbf{Q}(\mathbf{h}) = \bigotimes_{l=1}^u \mathbf{Q}_l(h_l), \text{ com } \mathbf{h} \in \Gamma, \quad (4.88)$$

em que

$$\begin{cases} \mathbf{Q}_l(0) = \frac{1}{a_l} \mathbf{J}_{a_l}, & \text{com } l = 1, \dots, u \\ \mathbf{Q}_l(1) = \mathbf{I}_{a_l} - \frac{1}{a_l} \mathbf{J}_{a_l}, & \text{com } l = 1, \dots, u \end{cases} \quad (4.89)$$

vê-se facilmente que  $\mathbf{Q}(\mathbf{h})$ , com  $\mathbf{h} \in \Gamma$ , são matrizes de projecção ortogonal, mutuamente ortogonais e constituem a base principal dessa Álgebra de Jordan Comutativa de matrizes simétricas, a  $\mathcal{A}(\Gamma)$ .

Como, ver [Fonseca *et al.*, 2003]

$$\mathbf{M}(\mathbf{h}) = \sum_{\mathbf{k} \in \Gamma(\mathbf{h})} b(\mathbf{k}) \mathbf{Q}(\mathbf{k}), \quad (4.90)$$

com

$$\Gamma(\mathbf{h}) = \{\mathbf{k} : \mathbf{k} \in \Gamma, k_l \leq h_l, l = 1, \dots, u\} \quad (4.91)$$

e

$$b(\mathbf{k}) = \prod_{l \notin \varphi(\mathbf{k})} a_l \quad (4.92)$$

onde

$$\varphi(\mathbf{k}) = \{l : k_l = 1\} \quad (4.93)$$

tem-se  $\mathbf{M}(\mathbf{h}) \in \mathcal{A}(\Gamma)$ , com  $\mathbf{h} \in \Gamma$ , já que

$$\mathbf{I}_a = \left( \mathbf{I}_a - \frac{1}{a} \mathbf{J}_a \right) + \frac{1}{a} \mathbf{J}_a. \quad (4.94)$$

Note-se que  $\mathbf{M}(\mathbf{h})$  será uma matriz que é combinação linear das matrizes da base principal da Álgebra de Jordan.

Por outro lado, tem-se

$$\mathbf{Q}(\mathbf{h}) = (\mathbf{A}(\mathbf{h}))^t \mathbf{A}(\mathbf{h}), \text{ com } \mathbf{h} \in \Gamma \quad (4.95)$$

e

$$\mathbf{A}(\mathbf{h}) = \bigotimes_{l=1}^u \mathbf{A}_l(h_l) \quad (4.96)$$

onde

$$\begin{cases} \mathbf{A}_l(0) = \frac{1}{\sqrt{a_l}} (\mathbf{1}^{a_l})^t, & \text{com } l = 1, \dots, u \\ \mathbf{A}_l(1) = \mathbf{L}_{a_l}, & \text{com } l = 1, \dots, u \end{cases} \quad (4.97)$$

e  $\mathbf{L}_{a_l}$ , com  $l = 1, \dots, u$  é obtida retirando a primeira linha igual a

$$\left[ \frac{1}{\sqrt{a_l}} \quad \dots \quad \frac{1}{\sqrt{a_l}} \right], \text{ com } l = 1, \dots, u \quad (4.98)$$

a uma matriz ortogonal  $a_l \times a_l$ , com  $l = 1, \dots, u$ .

Por exemplo, podemos ter

$$\mathbf{L}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},$$

$$\mathbf{L}_3 = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

e

$$\mathbf{L}_4 = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

já que as matrizes

$$\mathbf{P}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},$$

$$\mathbf{P}_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

e

$$\mathbf{P}_4 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

são ortogonais estandardizadas.

É agora fácil de aplicar os resultados da secção anterior, na medida em que construímos as matrizes  $\mathbf{A}_i$ , com  $i = 1, \dots, 2^u$ .

O que diferencia o nosso tratamento do habitual é:

- admitirmos ter  $\mathbf{Y}$  independente de  $S$ , com  $S \sim \sigma^2 \chi_g^2$ , o que permite eliminar restrições em  $\boldsymbol{\mu}$  para estimar  $\sigma^2$ ;
- admitir que  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{C})$ , sendo, em geral,  $\mathbf{C} \neq \mathbf{I}_u$ , mas conhecida, pelo que em vez de homocedasticidade teremos heterocedasticidade controlada.





# Capítulo 5

## Aplicação Prática

### 5.1 Considerações Prévias

A aplicação prática desta dissertação será apresentada no âmbito da metodologia *Statis* pelo que na secção seguinte iremos fazer o enquadramento dessa metodologia e de seguida apresentar-se-á a aplicação prática.

### 5.2 Metodologia *Statis*

A metodologia *Statis* foi inicialmente introduzida por Escoufier [Escoufier, 1973] e L'Hermier des Plantes, no Laboratório de Probabilidades e Estatística da Universidade de Montpellier II, por volta de 1976 [L' Hermier, 1976] e posteriormente desenvolvida por Lavit em 1988 [Lavit, 1988] e em 1994 [Lavit *et al.*, 1994].

Esta metodologia consiste no estudo simultâneo de várias tabelas de dados **quantitativos** ou **qualitativos transformados**, ver [Oliveira, 1996], que apresentem uma das seguintes características:

- $k$  tabelas de dados recolhidas em diferentes circunstâncias temporais ou espaciais, que apresentam os mesmos indivíduos, mas as variáveis podem eventualmente ser diferentes;
- $k$  tabelas de dados compostas pelas mesmas variáveis, mas onde o grupo de indivíduos presente em cada uma das tabelas pode, eventualmente, ser diferente.

Para cada uma destas situações corresponde uma estratégia: a primeira privilegia as posições relativas dos indivíduos (**Statis**), e a outra estuda as relações entre as variáveis (**Statis Dual**). Quando existe interesse em estudar o comportamento dos indivíduos e das variáveis, opta-se por aplicar as duas estratégias.

Existem muitos estudos desenvolvidos através da metodologia *Statis*, com aplicações em diferentes domínios do conhecimento. Contudo, **qualquer fenómeno**

cuja análise compreenda informação medida num horizonte temporal ou espacial, desde que seja constituída por dados quantitativos ou qualitativos transformados pode ser estudado por esta metodologia.

A aplicação da metodologia *Statis* envolve três fases :

1. **Inter-estrutura:** nesta fase pretende-se encontrar as relações existentes entre as diferentes matrizes de dados, de modo a verificar as semelhanças e as diferenças que surgem entre as matrizes de dados, efectuando-se assim uma comparação global das matrizes de dados;
2. **Compromisso:** esta fase consiste em resumir as  $k$  matrizes de dados numa só, através da definição de um compromisso, e em que este compromisso deve ser representativo do conjunto de tabelas e caracterizar a estrutura comum ao conjunto de dados;
3. **Intra-estrutura:** nesta fase consegue-se obter a representação das trajectórias que permitem visualizar a evolução dos indivíduos, bem como a quantificação dessa evolução.

Na figura 5.2, podemos ver as duas fases da metodologia *Statis*.

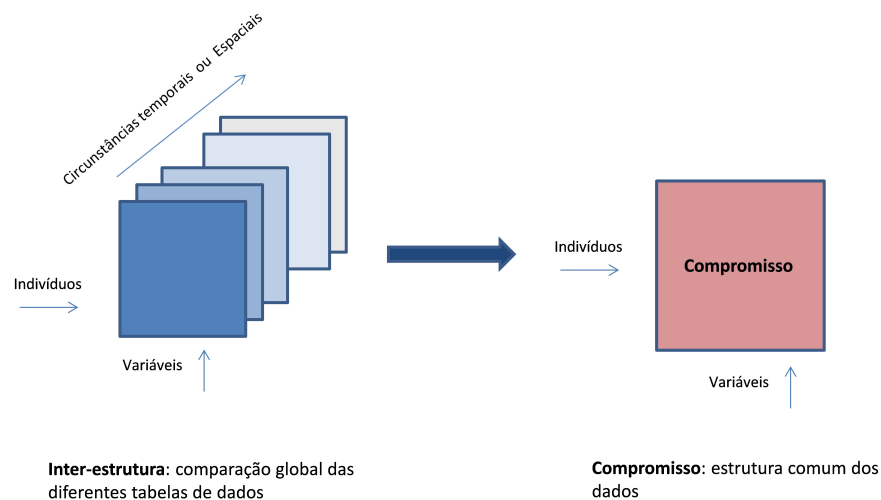


Figura 5.1: Fases da metodologia *STATIS*: Inter-estrutura e Compromisso.

Deste modo, podemos dizer que a metodologia *Statis*, sendo uma metodologia de Análise Conjunta de Tabelas, visa essencialmente captar as trajectórias dos indivíduos e das variáveis ao longo do tempo ou espaço, de modo a apresentar as semelhanças e as diferenças entre as diferentes tabelas de dados, ao longo do período ou espaço em estudo.

Assim, a metodologia *Statis* é uma **técnica de análise de dados exploratória**, inserida na análise conjunta de tabelas, **que se aplica a dados do tipo quantitativo ou qualitativo transformado** descritos por um conjunto de variáveis medidas em diferentes horizontes temporais ou espaciais. Os dados são representados em  $k$  tabelas designadas por  $X_i$ , com  $i = 1, \dots, k$ , em que  $i$  é o índice associado a cada matriz de dados que compõe o estudo. O objectivo principal desta metodologia consiste na procura de uma estrutura comum entre as tabelas de dados.

Assim, podemos dizer que esta metodologia não se restringe apenas à análise de uma matriz de dados, como seria o caso de uma Análise em Componentes Principais (ACP), mas sim, permitir a exploração simultânea de várias tabelas de dados quantitativos ou qualitativos transformados [Oliveira, 1996], usando o método *Statis* ou método *Statis-Dual*, conforme a situação.

### 5.2.1 O Método STATIS

Para aplicação do **método Statis**, em que o número de indivíduos é fixo podendo variar o número de variáveis, consideremos  $k$  matrizes de dados, cada uma delas denotada por  $\mathbf{X}_i$ , com  $i = 1, \dots, k$ , e do tipo  $n \times p_i$ , onde  $n$  é o número de indivíduos em cada tabela e  $p_i$  o número de variáveis (quantitativas ou qualitativas transformadas, ver [Oliveira, 1996]) na  $i$ -ésima tabela, com  $i = 1, \dots, k$ , como podemos ver através da figura 5.2

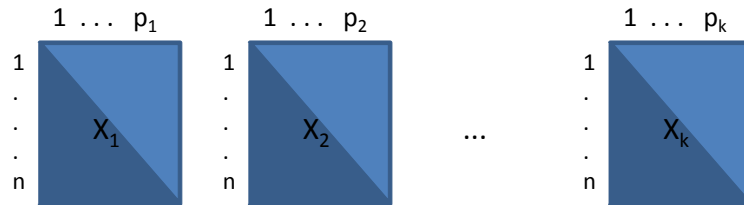


Figura 5.2:  $k$  quadros de dados: notação do método STATIS.

No instante  $i$ , o quadro de dados  $\mathbf{X}_i$  é a matriz de dimensão  $n \times p_i$

$$\mathbf{X}_i = \begin{bmatrix} (x_1^1)^i & (x_1^2)^i & \dots & (x_1^{p_i})^i \\ (x_2^1)^i & (x_2^2)^i & \dots & (x_2^{p_i})^i \\ \vdots & \vdots & \ddots & \vdots \\ (x_n^1)^i & (x_n^2)^i & \dots & (x_n^{p_i})^i \end{bmatrix}, \quad (5.1)$$

a  $j$ -ésima variável é o vector de  $\mathbb{R}^n$ , que contém as observações dessa variável

para cada um dos indivíduos,

$$(\mathbf{x}^j)^i = \begin{bmatrix} (x_1^j)^i \\ (x_2^j)^i \\ \vdots \\ (x_n^j)^i \end{bmatrix} \quad (5.2)$$

e o  $h$ -ésimo indivíduo é o vector de  $\mathbb{R}^{p_i}$  que contém as suas características para as  $p_i$  variáveis consideradas,

$$\left( (\mathbf{x}_h)^i \right)^t = \left[ (x_h^1)^i \quad (x_h^2)^i \quad \dots \quad (x_h^{p_i})^i \right]^t. \quad (5.3)$$

No âmbito do *Statis*, consideram-se  $k$  matrizes de dados, e a estas associam-se matrizes diagonais de pesos das variáveis e dos objectos. Os trios constituídos por estas matrizes será um estudo. Neste caso tem-se:

- as matrizes de dados  $\mathbf{X}_i$  do tipo  $n \times p_i$ , com  $i = 1, \dots, k$ ;
- as matrizes diagonais dos pesos dos objectos,  $\mathbf{D}_n$ ;
- as matrizes diagonais dos pesos das variáveis,  $\mathbf{D}_{p_i}$ , com  $i = 1, \dots, k$ .

O trio de matrizes  $(\mathbf{X}_i, \mathbf{D}_n, \mathbf{D}_{p_i})$  denomina-se por estudo e o conjunto dos vários estudos que constituem as várias observações designa-se por série de estudos.

O método *Statis* é um dos métodos de condensar a informação contida num trio de matrizes.

Assim, em [Y. Escoufier e H. L'Hermier, 1978], associou-se a cada estudo, as matrizes

$$\mathbf{A}_i = \mathbf{X}_i \mathbf{D}_{p_i} \mathbf{X}_i^t \mathbf{D}_n, \text{ com } i = 1, \dots, k, \quad (5.4)$$

de dimensão  $n \times n$  e interpretando-se

$$\mathbf{S} = [s_{ij}] = \langle \mathbf{A}_i, \mathbf{A}_j \rangle_{HS} = \text{tr} (\mathbf{A}_i \mathbf{A}_j^t) \quad (5.5)$$

como uma matriz de covariância.

### 5.2.2 O Método *Statis-Dual*

Para aplicação do **método** *Statis-Dual*, em que o número de indivíduos pode ser variável e o número de variáveis é fixo, consideremos  $k$  matrizes de dados, cada uma delas denotada por  $\mathbf{X}_i$ , com  $i = 1, \dots, k$ , e do tipo  $n_i \times p$ , onde  $n_i$  é o número de indivíduos na  $i$ -ésima tabela, com  $i = 1, \dots, k$  e  $p$  o número de

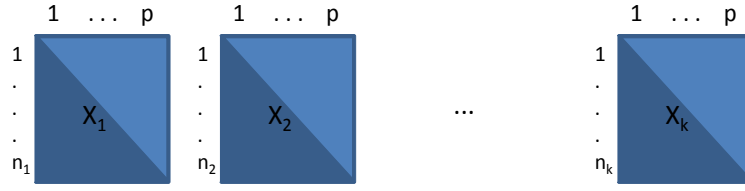


Figura 5.3:  $k$  quadros de dados: notação do método *STATIS-Dual*.

variáveis (quantitativas ou qualitativas transformadas, ver [Oliveira, 1996]) em cada tabela, como podemos ver através da figura 5.3

No instante  $i$ , o quadro de dados  $\mathbf{X}_i$  é a matriz de dimensão  $n_i \times p$

$$\mathbf{X}_i = \begin{bmatrix} (x_1^1)^i & (x_1^2)^i & \cdots & (x_1^p)^i \\ (x_2^1)^i & (x_2^2)^i & \cdots & (x_2^p)^i \\ \vdots & \vdots & \ddots & \vdots \\ (x_{n_i}^1)^i & (x_{n_i}^2)^i & \cdots & (x_{n_i}^p)^i \end{bmatrix}, \quad (5.6)$$

a  $j$ -ésima variável é o vector de  $\mathbb{R}^{n_i}$ , que contém as observações dessa variável para cada um dos indivíduos,

$$(\mathbf{x}^j)^i = \begin{bmatrix} (x_1^j)^i \\ (x_2^j)^i \\ \vdots \\ (x_{n_i}^j)^i \end{bmatrix} \quad (5.7)$$

e o  $h$ -ésimo indivíduo é o vector de  $\mathbb{R}^p$  que contém as suas características para as  $p$  variáveis consideradas,

$$\left( (\mathbf{x}_h)^i \right)^t = \left[ (x_h^1)^i \quad (x_h^2)^i \quad \cdots \quad (x_h^p)^i \right]^t. \quad (5.8)$$

No âmbito do *Statis-Dual*, consideram-se  $k$  matrizes de dados, e a estas associam-se matrizes diagonais de pesos das variáveis e dos objectos. Neste caso tem-se:

- as matrizes de dados  $\mathbf{X}_i$  do tipo  $n_i \times p$ , com  $i = 1, \dots, k$ ;
- as matrizes diagonais dos pesos dos objectos,  $\mathbf{D}_{n_i}$ , com  $i = 1, \dots, k$ ;
- as matrizes diagonais dos pesos das variáveis,  $\mathbf{D}_p$ .

O trio de matrizes  $(\mathbf{X}_i, \mathbf{D}_{n_i}, \mathbf{D}_p)$  denomina-se por estudo e o conjunto dos vários estudos que constituem as várias observações designa-se por série de estudos.

O método *Statis-Dual* é um dos métodos de condensar a informação contida num trio de matrizes.

Assim, em [Y. Escoufier e H. L'Hermier, 1978], associou-se a cada estudo, as matrizes

$$\mathbf{B}_i = \mathbf{X}_i^t \mathbf{D}_{n_i} \mathbf{X}_i \mathbf{D}_p, \text{ com } i = 1, \dots, k, \quad (5.9)$$

de dimensão  $p \times p$  e interpretando-se

$$\mathbf{S} = [s_{ij}] = \langle \mathbf{B}_i, \mathbf{B}_j \rangle_{HS} = \text{tr} (\mathbf{B}_i \mathbf{B}_j^t) \quad (5.10)$$

como uma matriz de covariância.

### 5.2.3 Inter-estrutura

Consideremos a matriz  $\mathbf{S} = [s_{ij}]$ , com  $i, j = 1, \dots, k$ , (matriz de inter-estrutura, simétrica, semidefinida positiva) designada por matriz dos produtos escalares entre as matrizes  $\mathbf{A}_i$  e  $\mathbf{A}_j$  ou entre as matrizes  $\mathbf{B}_i$  e  $\mathbf{B}_j$  conforme seja o método *Statis* ou o método *Statis-Dual*. Sem perda desta informação, seja  $\mathbf{S} = [s_{ij}]$ , com  $i, j = 1, \dots, k$ , uma matriz de dimensão  $k \times k$ , cujo termo geral é

$$s_{ij} = \langle \mathbf{A}_i, \mathbf{A}_j \rangle_{HS} = \text{tr} (\mathbf{A}_i \mathbf{A}_j^t), \quad (5.11)$$

e sendo  $\mathbf{S}$  uma matriz simétrica, terá como decomposição espectral

$$\mathbf{S} = \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t = \sum_{i=1}^k \sqrt{\theta_i} \gamma_i \sqrt{\theta_i} \gamma_i^t, \quad (5.12)$$

em que  $\theta_1 > \theta_2 > \dots > \theta_k$  são os valores próprios da matriz  $\mathbf{S}$  associados aos vectores próprios  $\gamma_1, \gamma_2, \dots, \gamma_k$ , respectivamente. Através da decomposição espectral da matriz  $\mathbf{S}$  é possível uma representação geométrica dos estudos. As coordenadas do ponto representativo do  $j$ -ésimo estudo são as  $j$ -ésimas componentes dos vectores  $\sqrt{\theta_1} \gamma_1, \sqrt{\theta_2} \gamma_2, \dots, \sqrt{\theta_k} \gamma_k$ , o que matricialmente se pode representar

$$\begin{bmatrix} \sqrt{\theta_1} \gamma_1 & \dots & \sqrt{\theta_j} \gamma_j & \dots & \sqrt{\theta_k} \gamma_k \end{bmatrix} = \begin{bmatrix} \sqrt{\theta_1} \gamma_{1,1} & \dots & \sqrt{\theta_2} \gamma_{2,1} & \dots & \sqrt{\theta_k} \gamma_{k,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sqrt{\theta_1} \gamma_{1,j} & \dots & \sqrt{\theta_2} \gamma_{2,j} & \dots & \sqrt{\theta_k} \gamma_{k,j} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sqrt{\theta_1} \gamma_{1,k} & \dots & \sqrt{\theta_2} \gamma_{2,k} & \dots & \sqrt{\theta_k} \gamma_{k,k} \end{bmatrix}, \quad (5.13)$$

ou seja, as coordenadas do ponto representativo do  $j$ -ésimo estudo estão representadas na  $j$ -ésima linha da matriz (5.13).

Após a representação geométrica dos pontos representativos dos estudos, será possível fazer a interpretação. Segundo a abordagem feita em [Lavit, 1988], Lavit considera que uma série de estudos tem uma estrutura comum, quando os pontos representativos dos vários estudos, segundo a representação de Escoufier, se situam ao longo de um único eixo, ou seja, as restantes componentes dos pontos serão praticamente nulas, à excepção da primeira, tendo-se então

$$\theta_j^2 = \|\theta_j \gamma_j\|^2 \approx 0, \text{ com } j = 2, \dots, k$$

e como consequência, a informação da série fica condensada numa parte determinística, afirmando-se assim que

$$\mathbf{S} \approx \theta_1 \gamma_1 \gamma_1^t.$$

Podemos então falar na existência de um colapso dimensional do conjunto dos pontos representativos da série de estudos. No caso referido em [Lavit, 1988], este colapso é unidimensional. Pode-se no entanto considerar colapsos  $r$ -dimensionais em que  $r < k$ . Dir-se-á então, que se está perante uma estrutura comum de grau  $r$ , ou seja, as últimas  $k - r$  coordenadas dos pontos representativos da série serão desprezáveis, o mesmo é dizer que

$$\theta_j^2 = \|\theta_j \gamma_j\|^2 \approx 0, \text{ com } j = r + 1, \dots, k.$$

## 5.3 Eleições Legislativas em Portugal Continental

### 5.3.1 Enquadramento da aplicação prática

Nesta secção pretende-se apresentar uma aplicação prática, evidenciando o objectivo proposto nesta dissertação, ou seja, pretende-se validar a escolha do grau do modelo.

Assim, o nosso estudo incidiu sobre os resultados de 12 anos de eleições legislativas realizadas em Portugal Continental entre os anos 1976 e 2009.

Para tal deveriam ser consideradas as NUTS (Nomenclaturas de Unidades Territoriais). As NUTS designam as sub-regiões estatísticas em que se divide o território dos países da União Europeia, incluindo o território português. Estas designações estão de acordo com o Regulamento (CE) n.º 1059/2003 do Parlamento Europeu e do Conselho de 26 de Maio de 2003, relativo à instituição de uma Nomenclatura Comum das Unidades Territoriais Estatísticas (NUTS), com vista a estabelecer uma divisão coerente e estruturada do território económico comunitário, criando uma base territorial comum para efeitos de análise estatística

de dados. Esta classificação é hierárquica, subdividindo cada Estado Membro em unidades territoriais ao nível de NUTS I, cada uma das quais é subdividida em unidades territoriais ao nível de NUTS II, sendo estas, por sua vez, subdivididas em unidades territoriais ao nível de NUTS III.

Assim, Portugal encontra-se dividido, de acordo com esta classificação

- **NUTS I** : Continente, Açores e Madeira
- **NUTS II** - Norte, Centro, Lisboa, Alentejo, Algarve, Açores e Madeira
- **NUTS III**
  - **Norte**: Minho-Lima, Cávado, Ave, Grande Porto, Tâmega, Entre Douro e Vouga, Douro e Alto Trás-os-Montes
  - **Centro**: Baixo Vouga, Baixo Mondego, Pinhal Litoral, Pinhal Interior Norte, Dão-Lafões, Pinhal Interior Sul, Serra da Estrela, Beira Interior Norte, Beira Interior Sul, Cova da Beira, Oeste e Médio Tejo
  - **Lisboa**: Grande Lisboa e Península de Setúbal
  - **Alentejo**: Alentejo Litoral, Alto Alentejo, Alentejo Central, Baixo Alentejo e Lezíria do Tejo
  - **Algarve**: Algarve
  - **Açores**: Açores
  - **Madeira**: Madeira

Uma vez que os dados que iremos apresentar dizem respeito a Portugal Continental e o primeiro ano a que dizem respeito os dados é 1976, iremos utilizar uma outra divisão do território português, por ser mais simples, pois a actual estrutura da divisão administrativa de Portugal é actualmente bastante complexa.

A Constituição Portuguesa de 1976, estabelece que Portugal se divide somente em regiões autónomas (Açores e Madeira) e em regiões administrativas. Por sua vez, todas estas se dividem em municípios e estes últimos em freguesias. As regiões administrativas, cuja criação está prevista desde a aprovação da Constituição (1976), são segundo este documento, os 18 distritos de Portugal Continental, pois “enquanto as regiões administrativas não estiverem concretamente instituídas, subsistirá a divisão distrital no espaço por elas não abrangido”, pelo que em Portugal Continental os municípios mantiveram-se agrupados em 18 distritos.

Assim, de acordo com a antiga divisão do território português, foram consideradas as seguintes regiões de Portugal Continental:

- **Região Norte** - que contempla os distritos de Viana do Castelo, Braga, Porto, Vila Real e Bragança



- **Região Centro Norte** - que contempla os distritos de Aveiro, Coimbra, Viseu, Guarda e Castelo Branco
- **Região Centro** - que contempla os distritos de Leiria, Lisboa, Santarém e Portalegre
- **Região Sul** - que contempla os distritos de Setúbal, Faro, Évora e Beja.

Em cada região foram seleccionados dois distritos: um distrito no Litoral e um distrito no Interior. Para cada distrito foram ainda considerados dois concelhos, sendo um deles a capital de distrito e o outro seleccionado aleatoriamente.

Na tabela 5.1 podemos ver os distritos escolhidos, bem como os concelhos analisados e respectivas abreviaturas.

Região	Distrito	Concelho
Norte	Braga	Braga (Brg)
		Fafe (Faf)
	Bragança	Alfândega da Fé (AlF)
		Bragança (Brgç)
Centro Norte	Coimbra	Coimbra (Cbr)
		Cantanhede (Ctn)
	Guarda	Guarda (Grd)
		Aguiar da Beira (AgB)
Centro	Leiria	Leiria (Ler)
		Óbidos (Obd)
	Santarém	Santarém (Str)
		Abrantes (Abr)
Sul	Setúbal	Setúbal (Stb)
		Barreiro (Brr)
	Beja	Beja (Bja)
		Almodovar (Almd)

Tabela 5.1: Distritos-Concelhos escolhidos por Região

Em cada concelho foram seleccionadas dez freguesias. No caso em que o concelho seleccionado não tinha dez freguesias, juntaram-se freguesias de concelhos adjacentes. Ao concelho de Óbidos juntámos três freguesias do concelho de Peniche, ao concelho de Abrantes juntámos três freguesias do concelho de Constância, ao concelho do Barreiro, juntámos três freguesias do concelho da Moita e três freguesias do concelho do Seixal, ao concelho de Setúbal juntámos quatro freguesias do concelho de Palmela e ao concelho de Almodôvar juntámos três freguesias do concelho de Mértola.

Os dados dos resultados eleitorais foram obtidos através do *site* do STAPE (*Secretariado Técnico dos Assuntos para o Processo Eleitoral*), onde para cada eleição apresentam o número de eleitores inscritos, o número de votantes, o

número de votos em branco, o número de votos nulos, o número de votos nos diferentes partidos e coligações e respectivos mandatos.

Com base na informação disponível, organizaram-se os dados de forma que para cada eleição e para cada concelho, fosse criada uma tabela de dupla entrada, onde para cada freguesia e para cada opção política apresenta-se as respectivas percentagens de votos. Na tabela 5.2, podemos ver as opções políticas que foram consideradas

OPÇÕES POLÍTICAS	
Partido Social Democrata	PSD
Partido Socialista	PS
Partido do Centro Democrático Social-Partido Popular	CDS-PP
Coligação Democrática Unitária	CDU
Outros Partidos de Direita	OutDRT
Outros Partidos de Esquerda	OutESQ
Votos em Brancos-Nulos	B-N
Abstenção	ABS

Tabela 5.2: As diferentes opções políticas consideradas

Em algumas eleições, o PSD e o CDS apresentaram listas conjuntas. Os votos nessa coligação foram distribuídos pelos dois partidos, tendo nestes casos sido ponderadas as suas votações proporcionalmente aos seus resultados nas eleições legislativas imediatamente anteriores ou posteriores.

### 5.3.2 Ajustamento e validação do modelo

Com base nos resultados eleitorais foi possível construir, para cada ano de eleição e para cada concelho uma matriz de dados  $\mathbf{X}_i$  com  $i = 1, \dots, 16$  com a percentagem de votos em cada uma das opções políticas consideradas para cada freguesia escolhida nesse concelho.

Assim, em cada matriz  $\mathbf{X}_i$  com  $i = 1, \dots, 16$ , as linhas da matriz correspondem às freguesias seleccionadas aleatoriamente e as colunas correspondem às opções políticas consideradas em 5.2.

Atribuímos as matrizes de pesos:  $\mathbf{D}_{10} = \frac{1}{10}\mathbf{I}_{10}$  (matriz de pesos para as freguesias) e  $\mathbf{D}_8 = \mathbf{I}_8$  (matriz dos pesos para as opções políticas).

Foi utilizado o *software* R (*versão 2.15.0 (2012-03-30)*, *The R Foundation for Statistical Computing, ISBN 3-900051-07-0*), para efectuar todos os cálculos necessários para o ajustamento e validação do modelo.

Após a obtenção da série de estudos  $(\mathbf{X}_i, \mathbf{D}_{10}, \mathbf{D}_8)$ , para cada concelho, recorreu-se ao *software* para obter as matrizes dos produtos escalares de Hilbert-Schmidt, bem como os valores e vectores próprios associados das matrizes  $\mathbf{S}$  das várias séries de estudos.

Os três primeiros valores próprios,  $\theta_1$ ,  $\theta_2$  e  $\theta_3$ , para os vários concelhos são apresentados na tabela 5.3.

<b>Concelho</b>	$\theta_1$	$\theta_2$	$\theta_3$
<b>Braga</b>	64443390.000	65229.950	50896.390
Fafe	83094160.324	305144.948	139302.726
<b>Bragança</b>	93442609.319	500487.791	171148.442
Alf. da Fé	78659763.114	373110.134	154047.373
<b>Coimbra</b>	74926310.000	115712.900	26936.890
Cantanhede	86888180.000	75238.370	29669.640
<b>Guarda</b>	75398565.017	540308.780	230426.971
Ag. da Beira	98230572.330	634681.620	284489.784
<b>Leiria</b>	95153290.000	248866.100	121078.800
Óbidos	78985540.000	271833.000	72716.980
<b>Santarém</b>	67324530.000	561305.500	104776.300
Abrantes	67144530.000	230306.800	143789.000
<b>Setúbal</b>	62339090.000	107597.900	22771.430
Barreiro	82344280.000	144983.100	17499.650
<b>Beja</b>	78848340.000	306586.200	122502.900
Almodôvar	83566161.553	359929.298	46847.421

Tabela 5.3: Valores próprios dos vários concelhos

Através da decomposição espectral das matrizes representativas dos vários estudos, apresenta-se na tabela 5.4, a percentagem de variância explicada, dada pelo quociente entre o primeiro valor próprio e o traço da matriz  $\mathbf{S}$ .

Observe-se que em todos os concelhos o peso do primeiro valor próprio na representação espectral é muito próximo dos 100% pelo que se confirma que a existir estrutura, esta será comum para as diferentes séries de estudos.

Note-se ainda, através da tabela 5.3, que em todos os concelhos, o primeiro valor próprio é muito superior aos restantes valores próprios, o que nos permite concluir que a estrutura comum tem grau um, ou seja,  $r = 1$ . Ora, [Oliveira e Mexia, 2004], mostraram que quando  $r = 1$  e  $k \leq 20$ , se

$$\tau_1 = \frac{\theta_1^2}{\sum_{j=2}^k \theta_j^2} \geq 200,$$

sendo  $\tau_1$  o valor da preponderância para  $\theta_1$ , podemos usar  $\tilde{\beta}_1 = \theta_1 \gamma_1$  para estimar  $\beta_1$ . De facto, neste caso temos  $r = 1$  e  $k = 12 \leq 20$ , e o valor da preponderância para cada concelho está apresentado na tabela 5.5.

Como podemos ver através da tabela 5.5, o valor da preponderância em todos os concelhos é superior a 13837, o que é um elevado valor para a preponderância,

<b>Concelho</b>	$\frac{\theta_1}{\sum_{i=1}^k \theta_i} \times 100\%$
<b>Braga</b>	99.68
Fafe	99.32
<b>Bragança</b>	99.12
Alf. da Fé	99.08
<b>Coimbra</b>	99.77
Cantanhede	99.82
<b>Guarda</b>	98.66
Ag. da Beira	98.71
<b>Leiria</b>	99.50
Óbidos	99.43
<b>Santarém</b>	98.92
Abrantes	99.25
<b>Setúbal</b>	99.77
Barreiro	99.78
<b>Beja</b>	99.36
Almodôvar	99.40

Tabela 5.4: Percentagem de variabilidade explicada nos vários concelhos

<b>Concelho</b>	<b>Valor da Preponderância, <math>\tau_1</math></b>
<b>Braga</b>	455165.930
Fafe	59512.878
<b>Bragança</b>	30549.259
Alf. da Fé	36228.008
<b>Coimbra</b>	392006.816
Cantanhede	1043315.325
<b>Guarda</b>	15794.299
Ag. da Beira	18753.902
<b>Leiria</b>	111704.967
Óbidos	74211.924
<b>Santarém</b>	13837.676
Abrantes	56569.224
<b>Setúbal</b>	319861.621
Barreiro	316138.465
<b>Beja</b>	56115.988
Almodôvar	52130.455

Tabela 5.5: Valor da Preponderância para os vários concelhos

garantindo-se assim a qualidade dos estimadores e permitindo concluir que a estrutura comum tem grau um.

Nas tabelas 5.6, 5.7, 5.8 e 5.9, indica-se, para cada concelho, o primeiro vector de estrutura, obtido pelo produto do primeiro valor próprio e as componentes do vector próprio associado, correspondendo a cada um dos doze anos eleitorais. Esta representação permite mais facilmente analisar se há maior ou menor variabilidade, uma vez que os vectores próprios são normalizados. Uma análise rápida das tabelas 5.6, 5.7, 5.8 e 5.9 permite-nos evidenciar a existência de uma maior variabilidade nas componentes do vector de estrutura para os concelhos de Leiria e Bragança sendo os concelhos de Beja e Setúbal os que apresentam maior estabilidade nas componentes do vector de estrutura.

Assim, nas tabelas 5.6, 5.7, 5.8 e 5.9 indica-se, para cada concelho, as componentes do vector estrutura, correspondendo as 12 componentes do vector estrutura aos 12 anos eleitorais. Este vector estrutura estimado, obteve-se fazendo o produto do valor próprio pelo vector próprio que lhe está associado, ou seja  $\tilde{\beta}_1 = \theta_1 \gamma_1$ .

	<b>Braga</b>	<b>Fafe</b>	<b>Bragança</b>	<b>Alf. Fé</b>
<b>1976</b>	16897578.60	18842922.97	19999578.25	20579019.43
<b>1979</b>	16282334.16	18264959.42	21263391.77	16686624.81
<b>1980</b>	15527366.28	19695811.61	22294824.84	18377587.85
<b>1983</b>	18311228.54	20435465.69	21312617.83	19341822.84
<b>1985</b>	13134725.81	18905203.72	22471853.78	18699778.67
<b>1987</b>	19468566.69	25591047.55	29861088.03	25334092.84
<b>1991</b>	21338046.03	28298001.09	30916941.76	25819204.87
<b>1995</b>	20699097.15	25979516.08	28346257.72	24869684.27
<b>1999</b>	20894689.86	26876356.16	30816876.51	25114586.00
<b>2002</b>	20232684.25	27161538.09	31018533.54	25983244.83
<b>2005</b>	19998804.21	26512399.37	30769988.61	24653804.26
<b>2009</b>	18565953.80	27557688.26	30223525.26	24150879.38

Tabela 5.6: Vectores de estrutura dos concelhos da Região Norte

Uma vez que concluímos que a estrutura comum tem grau um, apresentam-se os valores observados da estatística  $\overline{\mathcal{F}}$ , referida em (3.33) na página 53. Os valores dessa estatística e para os vários concelhos, encontram-se na tabela 5.10.

Como podemos observar pela tabela 5.10, os valores para a estatística  $\overline{\mathcal{F}}$  validam o modelo

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \alpha_i \alpha_i^t + \overline{\mathbf{E}}$$

para  $r = 1$ , de forma clara e para todos os concelhos, pois não permitem rejeitar a hipótese nula ( $H_0$  referida em (3.31) na página 53), a um nível de significância

	<b>Coimbra</b>	<b>Cantanhede</b>	<b>Guarda</b>	<b>Ag. Beira</b>
<b>1976</b>	21232561.01	21565223.89	18785757.84	26736782.96
<b>1979</b>	18562363.69	21804586.08	21477311.74	32412191.25
<b>1980</b>	18639341.10	22441098.71	18371004.44	25620905.25
<b>1983</b>	22656197.86	22914766.60	20224682.91	23774319.59
<b>1985</b>	15834455.43	20799719.56	16263256.01	21698386.88
<b>1987</b>	20533140.88	27667279.48	24089130.13	31899061.46
<b>1991</b>	22490150.20	28801653.14	20582292.12	30973982.79
<b>1995</b>	24649198.37	27054384.85	24714080.42	27494009.94
<b>1999</b>	24815774.41	26871054.18	25082674.35	28806125.17
<b>2002</b>	24295559.82	27145307.14	23810092.23	30829473.87
<b>2005</b>	22496849.82	25975552.14	25006886.79	28450197.56
<b>2009</b>	21445403.80	26213275.93	20582292.12	29476248.44

Tabela 5.7: Vectores de estrutura dos concelhos da Região Centro Norte

	<b>Leiria</b>	<b>Óbidos</b>	<b>Santarém</b>	<b>Abrantes</b>
<b>1976</b>	23076445.55	18207202.82	17923782.78	19012591.75
<b>1979</b>	25106921.77	16270935.96	15802530.23	13590164.57
<b>1980</b>	26410373.12	17432890.67	17269834.01	15759202.78
<b>1983</b>	22983465.59	20559979.04	18823265.86	18683434.41
<b>1985</b>	21926374.06	18168934.08	15018175.86	14693403.76
<b>1987</b>	37010334.08	24472054.89	19587144.33	18590510.70
<b>1991</b>	36249790.20	26646692.69	22660465.34	20216197.93
<b>1995</b>	27764792.04	24850274.0	21423716.23	21860160.05
<b>1999</b>	27167464.36	27051425.64	22267868.78	22230245.43
<b>2002</b>	29609046.71	26684669.12	21173998.90	24633253.21
<b>2005</b>	24343687.68	24592725.84	20829130.76	21801254.83
<b>2009</b>	22862141.57	24710814.79	18690223.17	18451383.15

Tabela 5.8: Vectores de estrutura dos concelhos da Região Centro

	<b>Setúbal</b>	<b>Barreiro</b>	<b>Beja</b>	<b>Almodovar</b>
<b>1976</b>	18031918.19	28923440.19	22214826.159	19823109.56
<b>1979</b>	16438877.04	28457675.90	26127815.551	20481953.42
<b>1980</b>	15459791.35	26693236.05	23202613.480	19218402.04
<b>1983</b>	17573278.25	26890446.90	22997041.365	22439101.98
<b>1985</b>	14105038.54	20666887.65	20001003.059	20559738.83
<b>1987</b>	16651781.70	20310862.98	20315354.628	23451962.09
<b>1991</b>	18416470.55	20704233.19	20662673.900	25605902.34
<b>1995</b>	19341095.93	22012907.41	22729499.185	27577880.94
<b>1999</b>	21004888.87	23598936.14	24534552.195	28779195.61
<b>2002</b>	20257990.72	22643791.80	23723624.881	27642788.99
<b>2005</b>	18791616.30	21072325.90	23924401.739	27613002.83
<b>2009</b>	18674991.01	20783953.35	21913748.603	23542175.91

Tabela 5.9: Vectores de estrutura dos concelhos da Região Sul

<u>Concelho</u>	<u>Valor da Estatística <math>\overline{\mathcal{F}}</math></u>
<b>Braga</b>	4.881
Fafe	2.879
<b>Bragança</b>	2.487
Alf. da Fé	3.252
<b>Coimbra</b>	2.137
Cantanhede	3.509
<b>Guarda</b>	3.016
Ag. da Beira	3.325
<b>Leiria</b>	2.934
Óbidos	2.485
<b>Santarém</b>	1.665
Abrantes	3.301
<b>Setúbal</b>	1.685
Barreiro	1.545
<b>Beja</b>	2.403
Almodôvar	1.945

Tabela 5.10: Valor da Estatística  $\overline{\mathcal{F}}$  para os vários concelhos

de 5%, o que leva à não rejeição de um modelo de grau um ( $r = 1$ ),

$$\mathbf{S} = \lambda_1 \boldsymbol{\alpha}_1 \boldsymbol{\alpha}_1^t + \bar{\mathbf{E}}$$

De facto, como podemos ver através da tabela 5.10, a hipótese de que os resíduos terem valores médios nulos e serem homocedásticos não foi rejeitada para nenhum dos concelhos. Com efeito, para um nível de significância de 5% e  $r = 1$ , a região de não rejeição para o teste com estatística  $\bar{\mathcal{F}}$  é  $\left[ \bar{f}_{\frac{q}{2}, 1, b_r - 1}; \bar{f}_{1 - \frac{q}{2}, 1, b_r - 1} \right] = \left[ \bar{f}_{0.025, 1, 65}; \bar{f}_{0.975, 1, 65} \right] = [0.00099; 5.265]$  para todos os concelhos.

Portanto, uma vez que os resíduos podem-se assumir com valores médios nulos e homocedásticos, foram calculados os valores da estatística  $\mathcal{F}_r$  para  $r = 1$ , apresentada em (3.28) na página 52, e que constam na tabela 5.11.

<u>Concelho</u>	<u>Valor da Estatística <math>\mathcal{F}_1</math></u>	<u>Estimativa de <math>\delta_1</math></u>
<b>Braga</b>	2503412.407	29130605.100
Fafe	327320.829	3808812.190
<b>Bragança</b>	168020.924	1955140.572
Alf. da Fé	199254.046	2318580.531
<b>Coimbra</b>	2156036.426	25088411.864
Cantanhede	5738235.047	66772177.640
<b>Guarda</b>	86868.643	1010823.116
Ag. da Beira	103146.458	1200237.697
<b>Leiria</b>	614377.383	7149106.642
Óbidos	408165.705	4749552.569
<b>Santarém</b>	76107.211	885599.178
Abrantes	311130.672	3620417.635
<b>Setúbal</b>	1759239.046	20471133.280
Barreiro	1738760.830	20232841.284
<b>Beja</b>	308637.957	3591411.499
Almodôvar	286717.503	3336337.126

Tabela 5.11: Valor da Estatística  $\mathcal{F}_1$  e Estimativa de  $\delta_1$  para os vários concelhos

Conforme podemos observar, os valores da estatística  $\mathcal{F}_1$  para os vários concelhos, foram altamente significativos, podendo pois não se rejeitar, para todos os concelhos, a existência de uma estrutura comum de grau um, ou seja, rejeitou-se a hipótese nula

$$H_{0,r} : \|\underline{\boldsymbol{\lambda}}^r\| = 0, \text{ para } r = 1,$$

conforme teste (3.42) da página 56.

Tendo em conta que se trata de uma estrutura comum de grau um, os valores da estatística  $\mathcal{F}_1^\circ$  (apresentada na proposição 3.2.3), para os vários concelhos, são os mesmos valores apresentados para a estatística  $\mathcal{F}_1$  e a distribuição destas



duas estatísticas coincide, nomeadamente, os parâmetros de não centralidade são iguais. Uma vez que os valores das estatísticas  $\mathcal{F}_1$  e  $\mathcal{F}_1^\circ$  são valores altamente significativos, não há dúvidas quanto à rejeição, para todos os concelhos, da hipótese nula

$$H_{0,1}^\circ : \lambda_1 = 0,$$

conforme teste (3.44) referido na página 57, reforçando assim a existência de uma estrutura comum de grau um.

Ainda em relação aos parâmetros de não centralidade, apresentam-se as estimativas para  $\delta_1$  na tabela 5.11, para os diferentes concelhos, obtidas através do estimador referido em (3.78) na página 66

$$\tilde{\delta}_r = (b_r - 2) \mathcal{T}_r^\circ - (k - r + 1),$$

sendo

$$\mathcal{T}_r^\circ = \frac{k - r + 1}{b_r} \mathcal{F}_r^\circ = \frac{\|\mathbf{v}_r (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t)\|^2}{\|b_{\text{vec}} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2} \sim \bar{F}_{k-r+1, b_r}(\delta_r)$$

e  $b_r = \frac{(k-r)(k-r+1)}{2}$ , para  $k = 12$  e  $r = 1$ .

Repare-se que não foi necessário construir intervalos de confiança para  $\delta_1$ , conforme apresentado na subsecção 3.2.3.4, uma vez que os valores das estatísticas  $\mathcal{F}_1$  e  $\mathcal{F}_1^\circ$  são valores altamente significativos, levando de imediato à rejeição da hipótese nula.

## 5.4 Semi-Simulação

Tornando-se necessário comprovar os resultados apresentados recorreremos a dados *semi-simulados*, utilizando para o efeito, os dados utilizados na secção 5.3.1 e procedendo a uma perturbação dos valores próprios das várias matrizes consideradas.

De facto, a utilização do termo dados *semi-simulados* deve-se ao facto de não estarmos interessados em gerar novos dados, efectuando uma simulação propriamente dita, mas sim obter novas matrizes  $\mathbf{S}$  tais que

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} \quad (5.14)$$

com grau  $r > 1$  e sendo  $\bar{\mathbf{E}} = \frac{1}{2} (\mathbf{E} + \mathbf{E}^t)$  tal que

$$\text{vec}(\mathbf{E}) \sim N\left(\mathbf{0}^{k^2}, \sigma^2 \mathbf{I}_{k^2}\right), \quad (5.15)$$

e cuja decomposição espectral de  $\mathbf{S}$  será

$$\mathbf{S} = \sum_{i=1}^k \theta_i \gamma_i \gamma_i^t.$$

Assim, uma vez que os resultados apresentados na secção 5.3.1, indicaram-nos uma estrutura comum com grau um ( $r = 1$ ), tornou-se necessário comprovar os resultados apresentados para  $r > 1$ .

De forma a conseguir que os dados apresentem um grau maior do que um, perturbou-se o 2º valor próprio das matrizes por forma que

$$\theta_2 = 0.25\theta_1,$$

mantendo-se os restantes valores próprios iguais aos considerados em 5.3.1.

De seguida, obtiveram-se as novas matrizes  $\mathbf{S}_i$ , com  $i = 1, \dots, 16$ , e para as quais se apresentam os seus quatro primeiros valores próprios,  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  e  $\theta_4$  na tabela 5.12.

<b>Matrizes</b>	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
<b>S<sub>1</sub></b>	64443390.000	16110850.000	50896.390	41678.700
<b>S<sub>2</sub></b>	83094160.324	20773540.081	139302.726	41575.630
<b>S<sub>3</sub></b>	93442609.319	23360652.330	171148.442	61161.520
<b>S<sub>4</sub></b>	78659763.114	19664940.779	154047.373	57289.282
<b>S<sub>5</sub></b>	74926310.000	18731580.000	26936.890	10562.830
<b>S<sub>6</sub></b>	86888180.000	21722050.000	29669.640	21014.170
<b>S<sub>7</sub></b>	75398565.017	18849641.254	230426.971	96759.961
<b>S<sub>8</sub></b>	98230572.330	24557643.083	284489.784	127329.927
<b>S<sub>9</sub></b>	95153290.000	23788320.000	121078.800	586396.620
<b>S<sub>10</sub></b>	78985540.000	19746390.000	72716.980	67040.760
<b>S<sub>11</sub></b>	67324530.000	16831130.000	104776.300	35840.490
<b>S<sub>12</sub></b>	67144530.000	16786130.000	143789.000	70442.250
<b>S<sub>13</sub></b>	62339090.000	15584770.000	22771.430	6734.857
<b>S<sub>14</sub></b>	82344280.000	20586070.000	17499.650	9255.178
<b>S<sub>15</sub></b>	78848340.000	19712080.000	122502.900	34092.210
<b>S<sub>16</sub></b>	83566161.553	20891540.388	46847.421	37877.930

Tabela 5.12: Valores próprios das várias matrizes *semi-simuladas*

Através da decomposição espectral das matrizes representativas dos vários estudos, apresenta-se na tabela 5.13 a percentagem de variância explicada dada pelo quociente entre o primeiro valor próprio e o traço da matriz  $\mathbf{S}_i$ , com  $i = 1, \dots, 16$ , bem como o quociente entre a soma dos dois primeiros valores próprios e o traço da respectiva matriz  $\mathbf{S}_i$ , com  $i = 1, \dots, 16$ . Note-se que através

<b>Matrizes</b>	$\frac{\theta_1}{\sum_{i=1}^k \theta_i} \times 100\%$	$\frac{\theta_1+\theta_2}{\sum_{i=1}^k \theta_i} \times 100\%$
<b>S<sub>1</sub></b>	79.86	99.83
<b>S<sub>2</sub></b>	79.80	99.75
<b>S<sub>3</sub></b>	79.77	99.72
<b>S<sub>4</sub></b>	79.71	99.64
<b>S<sub>5</sub></b>	79.95	99.94
<b>S<sub>6</sub></b>	79.44	99.93
<b>S<sub>7</sub></b>	79.59	99.49
<b>S<sub>8</sub></b>	79.58	99.47
<b>S<sub>9</sub></b>	79.84	99.81
<b>S<sub>10</sub></b>	79.85	99.82
<b>S<sub>11</sub></b>	79.84	99.79
<b>S<sub>12</sub></b>	79.74	99.67
<b>S<sub>13</sub></b>	79.96	99.96
<b>S<sub>14</sub></b>	79.97	99.96
<b>S<sub>15</sub></b>	79.83	99.79
<b>S<sub>16</sub></b>	79.89	99.86

Tabela 5.13: Percentagem de variabilidade explicada nas várias matrizes *semi-simuladas*

da tabela 5.13 podemos verificar que os dois primeiros valores próprios das matrizes  $\mathbf{S}_i$ , com  $i = 1, \dots, 16$ , são fortemente dominantes.

Para todas as matrizes  $\mathbf{S}_i$ , com  $i = 1, \dots, 16$ , o peso do primeiro valor próprio na representação espectral das matrizes  $\mathbf{S}_i$  é muito próximo dos 80%, e o peso da soma do primeiro e segundo valores próprios na representação espectral é muito próximo dos 100%, o que confirma que a existir estrutura, esta será comum para as diferentes séries de estudos, com possibilidade de ser uma estrutura comum de grau dois.

Os valores observados da estatística  $\overline{\mathcal{F}}$ , referida em 3.33 na página 53, considerando  $r = 2$ , para as várias matrizes  $\mathbf{S}_i$ , com  $i = 1, \dots, 16$ , encontram-se na tabela 5.14.

Os valores para a estatística  $\overline{\mathcal{F}}$  validam o modelo

$$\mathbf{S} = \sum_{i=1}^2 \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \overline{\mathbf{E}}$$

de forma clara e para todas as matrizes consideradas, pois não permitem rejeitar a hipótese nula ( $H_0$  referida em (3.31)), a um nível de significância de 5%, o que leva à não rejeição de um modelo de grau dois ( $r = 2$ ),

$$\mathbf{S} = \lambda_1 \boldsymbol{\alpha}_1 \boldsymbol{\alpha}_1^t + \lambda_2 \boldsymbol{\alpha}_2 \boldsymbol{\alpha}_2^t + \overline{\mathbf{E}}$$

Matrizes	Valor da Estatística $\overline{\mathcal{F}}$
$\mathbf{S}_1$	4.255
$\mathbf{S}_2$	3.181
$\mathbf{S}_3$	3.274
$\mathbf{S}_4$	4.348
$\mathbf{S}_5$	3.759
$\mathbf{S}_6$	4.442
$\mathbf{S}_7$	3.642
$\mathbf{S}_8$	3.999
$\mathbf{S}_9$	2.907
$\mathbf{S}_{10}$	3.327
$\mathbf{S}_{11}$	2.473
$\mathbf{S}_{12}$	2.912
$\mathbf{S}_{13}$	2.154
$\mathbf{S}_{14}$	3.203
$\mathbf{S}_{15}$	2.547
$\mathbf{S}_{16}$	5.270

Tabela 5.14: Valor da Estatística  $\overline{\mathcal{F}}$  para as várias matrizes *semi-simuladas*

com  $\overline{\mathbf{E}} = \frac{1}{2} (\mathbf{E} + \mathbf{E}^t)$  sendo

$$\text{vec}(\mathbf{E}) \sim N\left(\mathbf{0}^{12^2}, \sigma^2 \mathbf{I}_{12^2}\right). \quad (5.16)$$

De facto, como podemos ver através da tabela 5.14, a hipótese de que os resíduos terem valores médios nulos e serem homocedásticos não foi rejeitada para nenhuma das matrizes. Com efeito, para um nível de significância de 5% e  $r = 2$ , a região de não rejeição para o teste com estatística  $\overline{\mathcal{F}}$  é

$$\left[\overline{f}_{\frac{q}{2}, 1, b_r - 1}; \overline{f}_{1 - \frac{q}{2}, 1, b_r - 1}\right] = [\overline{f}_{0.025, 1, 54}; \overline{f}_{0.975, 1, 54}] = [0.000991; 5.3159]$$

para todas matrizes  $\mathbf{S}_i$ , com  $i = 1, \dots, 16$ .

Portanto, uma vez que os resíduos podem-se assumir com valores médios nulos e homocedásticos, foram calculados os valores da estatística  $\mathcal{F}_r$  para  $r = 2$ , apresentada em (3.28) na página 52, e que constam na tabela 5.15.

De facto, como podemos ver através da tabela 5.15, a hipótese nula

$$H_{0,r} : \|\underline{\boldsymbol{\lambda}}^r\| = 0, \text{ para } r = 2,$$

conforme teste (3.42) da página 56, foi rejeitada para todas as matrizes, pois os valores da estatística  $\mathcal{F}_2$  para as várias matrizes, foram altamente significativos, podendo pois não se rejeitar, para todas as matrizes  $\mathbf{S}_i$ , com  $i = 1, \dots, 16$ , a existência de uma estrutura comum de grau dois ( $r = 2$ ).

Matrizes	Valor da Estatística $\mathcal{F}_2$	Valor da Estatística $\mathcal{F}_2^\circ$
$\mathbf{S}_1$	2167068.706	266537.862
$\mathbf{S}_2$	765877.184	94198.798
$\mathbf{S}_3$	627932.826	77232.380
$\mathbf{S}_4$	497830.916	61230.540
$\mathbf{S}_5$	15311279.112	1883205.452
$\mathbf{S}_6$	12176418.667	1497634.381
$\mathbf{S}_7$	212404.043	26124.561
$\mathbf{S}_8$	219487.311	26995.766
$\mathbf{S}_9$	1203173.716	147983.933
$\mathbf{S}_{10}$	1558147.154	191643.768
$\mathbf{S}_{11}$	921992.332	113400.126
$\mathbf{S}_{12}$	429730.545	52854.559
$\mathbf{S}_{13}$	17255801.129	2122371.261
$\mathbf{S}_{14}$	40248585.374	4950360.768
$\mathbf{S}_{15}$	940556.884	115683.467
$\mathbf{S}_{16}$	4024130.126	494946.486

Tabela 5.15: Valor da Estatística  $\mathcal{F}_2$  e  $\mathcal{F}_1^\circ$  para as várias matrizes *semi-simuladas*

Tendo em conta que se trata de uma estrutura comum de grau dois, apresentam-se ainda na tabela 5.15 os valores da estatística  $\mathcal{F}_2^\circ$  (apresentada na proposição 3.2.3), para as várias matrizes. Uma vez que os valores da estatística  $\mathcal{F}_2^\circ$  são valores altamente significativos, não há dúvidas quanto à rejeição, para todos as matrizes, da hipótese nula

$$H_{0,2}^\circ : \lambda_2 = 0,$$

conforme teste (3.44) referido na página 57, reforçando assim a existência de uma estrutura comum de grau dois.

Ainda em relação aos parâmetros de não centralidade, apresentam-se as estimativas para  $\delta_2$  na tabela 5.16, para as diferentes matrizes, obtidas através do estimador referido em (3.78) na página 66

$$\tilde{\delta}_r = (b_r - 2) \mathcal{T}_r^\circ - (k - r + 1),$$

sendo

$$\mathcal{T}_r^\circ = \frac{k - r + 1}{b_r} \mathcal{F}_r^\circ = \frac{\|\mathbf{v}_r (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t)\|^2}{\|b\_ \text{vec} (\mathbf{P}^* \mathbf{S} (\mathbf{P}^*)^t ; r)\|^2} \sim \bar{F}_{k-r+1, b_r}(\delta_r)$$

e  $b_r = \frac{(k-r)(k-r+1)}{2}$ , para  $k = 12$  e  $r = 2$ .

<b>Matrizes</b>	<b>Estimativa de <math>\delta_2</math></b>
<b>S<sub>1</sub></b>	2825290.339
<b>S<sub>2</sub></b>	998496.260
<b>S<sub>3</sub></b>	818652.225
<b>S<sub>4</sub></b>	649032.729
<b>S<sub>5</sub></b>	19961966.794
<b>S<sub>6</sub></b>	15874913.444
<b>S<sub>7</sub></b>	276909.351
<b>S<sub>8</sub></b>	286144.115
<b>S<sub>9</sub></b>	1568618.690
<b>S<sub>10</sub></b>	2031412.937
<b>S<sub>11</sub></b>	1202030.340
<b>S<sub>12</sub></b>	560247.325
<b>S<sub>13</sub></b>	22497124.376
<b>S<sub>14</sub></b>	52473813.139
<b>S<sub>15</sub></b>	1226233.750
<b>S<sub>16</sub></b>	5246421.752

Tabela 5.16: Estimativa de  $\delta_2$  para as várias matrizes *semi-simuladas*

Repare-se, que mais uma vez, não foi necessário construir intervalos de confiança para  $\delta_2$ , conforme apresentado na subsecção 3.2.3.4, uma vez que os valores das estatísticas  $\mathcal{F}_2$  e  $\mathcal{F}_2^\circ$  são valores altamente significativos, levando de imediato à rejeição da hipótese nula.

# Capítulo 6

## Conclusão

Nesta dissertação mostramos como utilizar os operadores do tipo **vec** na validação do modelo. Recorde-se que os operadores **vec** são operadores que quando aplicados a uma matriz, empilham sucessivamente as colunas dessa matriz, ver [Harville, 1998] e [Schott, 1997]. Os operadores do tipo **vec** são operadores que também fazem corresponder vectores a matrizes, ou seja, permitem reorganizar alguns elementos de uma matriz em um vector coluna. A formulação apresentada permitiu-nos fazer inferência sobre as séries de estudo, pois os resultados apresentados nesta dissertação podem ser aplicados às matrizes dos produtos de Hilbert-Schmidt, matrizes estas muito importantes na primeira fase da metodologia *Statis*. Assim, os operadores permitiram-nos apresentar resultados que permitem realizar inferência para modelos para matrizes estocásticas simétricas. Estes modelos deram a base para realizar inferência para matrizes isoladas e famílias estruturadas de matrizes. Em particular consideramos o caso em que as matrizes correspondem aos tratamentos de modelos base.

Assim, usando os resultados apresentados, com a utilização de operadores do tipo **vec**, é possível ajustar um modelo do tipo

$$\mathbf{S} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} = \sum_{i=1}^r \beta_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}}$$

com grau  $r$ . Caso  $\lambda_r = 0$ , o modelo anterior poderá ser simplificado, obtendo-se

$$\mathbf{S} = \sum_{i=1}^{r-1} \lambda_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}} = \sum_{i=1}^{r-1} \beta_i \boldsymbol{\alpha}_i^t + \bar{\mathbf{E}}$$

se  $\lambda_{r-1}$  for significativamente diferente de 0. Neste caso, estaremos em presença de um modelo com grau  $r - 1$ . Estes resultados, como já referimos, podem ser aplicados às matrizes dos produtos de Hilbert-Schmidt, as quais são muito importantes na primeira fase da Metodologia *Statis*.

Em termos práticos, usamos os resultados das Eleições Legislativas que ocorreram entre 1976 e 2009 em Portugal Continental. Com estes dados obtivemos

um modelo com grau um. Sendo de nosso interesse que o grau fosse maior que um, recorremos a dados *semi-simulados*, ou seja, utilizamos os resultados dessas mesma Eleições Legislativas, mas em que perturbamos o segundo valor próprio, de forma a que este fosse fortemente dominante, como acontecia com o primeiro valor próprio. Assim, apresentamos um modelo para séries de estudo com estrutura comum de grau dois ( $r = 2$ ), e que através da validação deste modelo, verificou-se que não se rejeitava a hipótese de os resíduos terem valores médios nulos e serem homocedásticos e que a rejeição das hipóteses  $H_{0,2}$  e  $H_{0,2}^c$  indicava um ajustamento adequado de um modelo com estrutura comum de grau dois ( $r = 2$ ).

De futuro, pretendemos estudar os vários tratamentos, no caso em que o delinqueamento não tem estrutura ortogonal. Um outro aspecto a trabalhar, relaciona-se com a parte da simulação.

Como podemos constatar, existe um trabalho que nos espera no futuro, pois encaramos uma tese de doutoramento como sendo um trabalho que terá sempre continuação, e a partir da mesma, seja possível desenvolver um trabalho cada vez mais completo e pormenorizado.



# Bibliografia

Areia A. ; *Séries emparelhadas de estudo*. Tese de Doutoramento, Universidade de Évora; 2009.

Drygas, H. ; Zmyslony, R. ; *Jordan algebras and Bayesian quadratic estimation of variance components*. Linear Algebra and its Applications, 168:259-275; 1992.

Escoufier, Y. ; *Le traitement des variables vectorielles*. Biometrics, 29, pp. 750-760; 1973.

Escoufier, Y. ; L'Hermier, H. ; *A propos de la Comparasion Graphique des Matrices de Variance*. Biom. J., Vol. 20, N°5, pp. 477-483; 1978.

Ferreira, D. ; *Variáveis pivot indutoras e componentes de variância em modelos normais ortogonais*. Dissertação de Doutoramento em Matemática Estatística. Universidade da Beira Interior; 2005.

Ferreira, S. ; *Inferência para modelos ortogonais com segregação*. Dissertação de Doutoramento em Matemática Estatística. Universidade da Beira Interior; 2006.

Fonseca, M.; Mexia, J. T.; Zmyslony, R. ; *Estimators and tests for variance components in cross nested orthogonal models design*. Discussiones Mathematicae, Probab. Stat., 23 (2): 175–201; 2003.

Fonseca, M.; Mexia, J. T.; Zmyslony, R. ; *Binary operations on Jordan algebras and orthogonal normal models*. Linear Algebra and Its Applications, 411, 75–86; 2006.

Fonseca, M.; Mexia, J. T.; Zmyslony, R. ; *Jordan algebras, generating pivot variables and orthogonal normal models*. Journal of Interdisciplinary Mathematics, 1, 305–326; 2007.

Fonseca, M.; Mexia, J. T.; Zmyslony, R. ; *Inference in normal models with commutative orthogonal block structure*. Acta et Commentationes Universitatis Tartuensis de Mathematica, 12, 3–16; 2008.

- Graham, A. ; *Kronecker Products and Matrix Calculus with Applications*. Halsted Press, John Wiley and Sons, New York; 1981.
- Harville, A. ; *Matrix Algebra from a statistician's perspective*. Springer-Verlag; 1998.
- Horn, R. ; Johnson, C. ; *Matrix Analysis*. Cambridge University Press; 1985.
- Ito, K. ; *Robustness of Anova and Macanova Test Procedures*. P. R. Krishnaiah (ed), Handbook of Statistics, Vol. I. North Holland; 1980.
- Jesus, V., Fonseca, M., Mexia, J. T., Zmyslony, R. ; *Binary operations and canonical forms for factorial and related models*. Linear Algebra and its Applications, 43, 2781–2797; 2009.
- Jesus, V., Mexia, J. T., Rodrigues, P. C. ; *Binary operations on orthogonal models - application to prime basic factorials and fractional replicates*. Journal of Statistical Theory and Practice, 3 , 505–521; 2009.
- Jordan, P. ; Von Neumann, J. ; Wigner, E. ; *On an algebraic generalization of the quantum mechanical formulation*. Ann. Math. 36:26-64; 1934.
- Khuri, A. I. ; Mathew, B. ; Sinha, B. K. *Statistical Tests for Mixed Linear Models*. J. Wiley & Sons, New York; 1998.
- L'Hermier, H. ; *Structuration des tableaux a trois indices de la statisque*. Thèse de 3ème cycle. Université de Montpellier; 1976.
- Lavit, Ch. ; *Analyse Conjointe de Tableaux quantitatifs*. Collection Méthodes + Programmes, Masson, Paris; 1988.
- Lavit, Ch. ; Escoufier, Y. ; Sabatier, R. ; Traissac, P. ; *The ACT (Statis Method)*. Computacional Statistics and Data Analysis. Vol. 18, pp. 97-119; 1994.
- Loève, M; *Probability Theory*. 4th ed. ; New York; Springer-Verlag; 1977.
- Mexia, J. T.; *Standardized Orthogonal Matrices and the Decomposition of the Sum of Squares for Treatments*. Trabalhos de Investigação Nº1. Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa; 1988.
- Mexia J. T.; *Controlled Heteroscedasticity, Quotient Vector Spaces and F tests for Hypotheses on Mean Vectors*. Trabalhos de Investigação, Nº 2; Departamento de Matemática - Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa; 1989a.

- Mexia J. T.; *Simultaneous Confidence Intervals: Generalization of the Scheffé Theorem*. *Trabalhos de Investigação*, Nº2; Departamento de Matemática- Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa; 1989b.
- Mexia J. T.; *Best Linear Unbiased Estimates, Duality of F Tests and the Scheffé Multiple Comparison Method in Presence of Controlled Heterocedasticity*; *Computation Statistics & Data Analysis* - Vol. 10, Nº3 ; 271-281 ; 1990.
- Mexia, J. T.; *Introdução à inferência estatística linear*; Centro de Estudos de Matemática Aplicada -Edições Universitárias Lusófonas; Lisboa; 1995.
- Mexia, J.T.; Oliveira, M.M. ; *Asymptotic linearity and limit distributions, approximations*. *Journal of Statistical Planning and Inference*, Vol. 140, (2), 353-357; 2010.
- Mexia, J. T.; Nunes, C. ; Oliveira, M.M. ; *Multivariate application domains for the Delta method*. 9th International Conference on Numerical Analysis and Applied Mathematics. AIP Conf. Proc. Vol. 1389, 1486-1489; 2011.
- Mood, A. ; Graybill, R. ; Boes D. ; *Introduction to the Theory of Statistics*. Third Edition. McGRAW-HILL; 1987.
- Nunes, C. ; Oliveira, M. M. and Mexia, J. T. ; *Application domains for the delta method*. *Statistics*. Vol. 47 (2), 317-328 (12) ; 2013.
- Oliveira, M. M. ; *Análise dos Resultados Eleitorais e Análise da Evolução do Grau de Desfoliação dos Sobreiros (ACT)-Método STATIS*. Tese de Mestrado. Faculdade de Ciências da Universidade de Lisboa; 1996.
- Oliveira, M. M.; Mexia, J. ; *Tests for the rank of Hilbert-Schmidt product matrices*. *Advances in Data Science and classifications*. Rizzi, Vichi and Bock, Ed. Springer; 1998.
- Oliveira, M. M.; Mexia, J. ; *F tests for hypothesis on the structure vectors of series*. *Discussiones Mathematicae*, Vol.19, (2), pp. 345-353; 1999.
- Oliveira, M. M.; *Modelação de Séries Emparelhadas de Estudos com Estrutura Comum*. Tese de Doutoramento. Universidade de Évora; 2001.
- Oliveira, M. M.; Mexia, J. T. ; *AIDS in Portugal: Endemic versus Epidemic Forecasting Scenarios for Mortality*. *International Journal of Forecasting*, 20: 131-137; 2004.
- Pollock, D.S.G. ; *The Algebra of Econometrics*. John Wiley & Sons; 1979.
- Rao, C. R. ; *Linear Statistical Inference and Its Applications*. Second Edition; John Willey & Sons; New York; 1973.

- Rao, C. ; Kleffe, J. ; *Estimation of variance components and applications*. North Holland, Elsevier - Amsterdam; 1988.
- Salvador, D. ; Monteiro, S. ; Mexia, J.T.; *Structured families of stochastic symmetric matrices*. Proceedings of the 10th International Conference on Numerical Analysis and Applied Mathematics, Kos, Greece, September 19 – 25. AIP Conf. Proc. Vol. 1479, pp. 1698-1700; 2012.
- Salvador, D. ; Monteiro, Sandra I. C; Mexia, J. T.;. *Álgebras de Jordan Comutativas e Modelos Ortogonais*. 6th Workshop on Statistics, Mathematics and Computation and 3rd Portuguese-Polish Workshop on Biometry, Covilhã; 2012a.
- Salvador, D. ; Mexia, J. T.; *Vec type operators and adjusting models for symmetric stochastic matrices*. 7th Workshop on Statistics, Mathematics and Computation and 5th International Conference on Risk Analysis, Tomar; 2013.
- Scheffé, H. *The Analysis of Variance*. John Wiley & Sons. New York; 1959.
- Schott, J. R. ; *Matrix Analysis for Statistics*; Wiley - Interscience; 1997.
- Seely, J. ; *Linear spaces and unbiased estimation*. Ann. Math. Stat., 41, 1725-1734; 1970.
- Seely, J. ; *Quadratic Subspaces and Completeness*. Ann. Math. Stat. ; 42(2): 710-721; 1971.
- Seely, J. ; *Completeness for a family of multivariate normal distribution*. Ann. Math. Stat. 43:1644-1647; 1972.
- Seely, J. ; *Minimal sufficient statistics and completeness for multivariate normal families*. Sankyã 39:170-185; 1977.
- Steeb, W. H. ; *Kronecker Product of Matrices and Applications*. Mannheim, Germany - Bibliographisches Institut; 1991.
- Steeb, W. H. ; Hardy, Y. ; *Matrix calculus and Kronecker Product: a practical approach to linear and multilinear algebra*. Second Edition. World Scientific Publishing Company; 2011.
- Vanaleuween, D. ; Seely, J. ; Birkes, D. ; *Sufficient conditions for orthogonal designs in mixed linear models*. Journal of Statistical Planning and Inference 73, 373-389; 1998.
- Vanaleuween, D. ; Birkes, D. ; Seely, J. ; *Balance and orthogonality in designs for mixed classification models*. The Annals of Statistics 27-6, 1927-1947; 1999.
- Vaquinhas, R. ; Mexia, J. ; *Convergence of Matrices and Subspaces with Statistical Applications*. Distância 92-Rennes; 1992.

Zmyślony, R. ; *A characterization of best linear unbiased estimators in the general linear model*. Lectures Notes in Statistics 2; 1978.

.